CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centres of the CGIAR

SCIENCE COUNCIL SECRETARIAT

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

This document comprises:

(a) Forward by Hans Gregersen, Chair of the CGIAR Science Council, Standing Panel on Impact Assessment (SPIA)

(b) Report of the study - Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centres of the CGIAR

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centres of the CGIAR

SCIENCE COUNCIL SECRETARIAT

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

FOREWORD BY THE CGIAR SCIENCE COUNCIL, STANDING PANEL ON IMPACT ASSESSMENT

Since establishment in 1971, the CGIAR community has invested approximately seven billion dollars in various research and research related activities. In an era characterized by scarce development resources, it is relevant to ask: *Do the documented benefits from CGIAR research justify the total investment in the CGIAR so far?* Although the CGIAR System has been a world leader in documenting research impacts, no previous study has attempted to comprehensively address this question in a quantitative manner. Thus, this study, which has received strong support from a number of stakeholders, represents a first attempt to scale-up quantified economic impacts to a System level.

At various times, the overall efficacy of agricultural research as development assistance has been called into question. With this in mind, the present analysis is intended to resolve on a preliminary basis whether the *entire* investment in the CGIAR over time can be justified on the basis of the benefits derived from its proven (and agreed-upon) major successes. Prior impact analyses have been unable to directly address this issue, because such have focused on the costs and benefits only of research successes, while ignoring the costs of associated efforts that have not resulted in quantifiable impacts. The present analysis overcomes these constraints by compiling reliable estimates of large-scale benefits, and comparing such with the ***total*** investment in the System to-date, under a number of different explicitly stated assumptions. The reasoning is that if the accumulated, aggregate value of generally accepted and credible benefits from a group of CGIAR activities is at least equal to the value of the entire investment in the CGIAR, when an acceptable alternative rate of return to investment is used to discount/compound benefits and costs, then the investment is justified under the assumption that the sum of benefits from all other CGIAR projects is zero or positive.

SPIA retained a consultant, David Raitzer, to develop this preliminary assessment, which also serves as a basis for deciding whether a more detailed, exhaustive assessment would be worth undertaking in the future. SPIA recognized from the beginning that it was impossible to carry out a complete benefit-cost analysis (comparing total returns with total investment), since the benefits derived from many of the CGIAR outputs have not been measured and, indeed, probably cannot easily be valued in economic terms because: (a) most research monitoring systems are not set up with that need in mind; (b) many important impacts are difficult to attribute; and/or (c) many desirable effects are difficult to value in market price terms. Thus, while aggregate investment can be estimated with some confidence, a comprehensive aggregate benefit figure cannot be easily derived. Consequently, the approach SPIA followed is inherently conservative, and provides only partial estimates of the benefits that have accrued beyond the breakeven point, at different levels of confidence. In fact, it is widely accepted that there are other positive benefits from the rest of the CGIAR investment, even if not yet documented.

More specifically, the present approach involved: (a) identifying available economic impact assessments (IAs) of CGIAR investments showing significant net benefits, (b) synthesising the methodological literature into standards for ex-post impact assessment ‘plausibility’ (c) appraising the transparency and analytical rigour of the benefit estimates provided by identified studies; and (d) adding up the benefits from those studies that met certain standards of rigour, starting with the most highly credible group of benefit estimates, followed by more inclusive standards to see what the relationship was between the entire seven billion dollar investment and the benefits generated at each chosen level of plausibility.

This type of study does face a number of limitations. Such an approach does not allow for appraisal of the rates of return on incremental investments in the CGIAR, nor the comparative efficacy of individual programs or projects. However, the main issue addressed, the demonstrated efficacy of the aggregate CGIAR investment, is critical both for the future of the CGIAR, and for the future of agricultural research as development assistance. In addition, such analysis can help to highlight those areas of research that are proven sources of benefits, and can help identify critical gaps in the CGIAR's IA portfolio.

To arrive at a collective credible minimum benefit figure, Raitzer spent much time and effort developing a hierarchical framework of principles, criteria and indicators for judging whether benefit estimates from individual studies are 'plausible' and could with confidence be included in the aggregate benefit values. These standards alone comprise an important output of the analysis, and will be built upon in the future as a tool for Centres and investors to use in designing and screening impact studies.

With review standards applied, aggregate benefit estimates for different levels of confidence were developed. The study concludes that even under the most conservative of the six plausibility scenarios, the widely-accepted benefits from a few highly successful CGIAR programmes easily justify the entire aggregate investment in the CGIAR. Although the overall result may not be surprising, since other major assessments of the CGIAR and some of its large-scale investments provide evidence of the high returns to CGIAR research, the holistic nature of these findings is novel and persuasive.[1] This study represents the first attempt to aggregate the documented economic benefits produced by the System, and, as such, it offers compelling evidence that the System's investment has been a productive one.

While, in aggregate, the evidence is impressive, this study does identify a number of ways in which the persuasiveness of individual studies could be further enhanced. In particular, topical coverage by large-scale IAs is somewhat limited, and counterfactual development could benefit from additional attention. In addition, the present analysis notes that increased transparency would strengthen the confidence of results, and more reliable data sources would enhance precision.

One particularly interesting aspect of the aggregate benefit-cost ratios estimated relates to the sensitivity of results to simple assumptions applied in the analysis. For example, a relatively minor slackening of the standards required for demonstration of impact increases the ratio of benefits to costs from under 2 to over 17. Such susceptibility to minor variances in study standards underscores the need for greater consensus regarding minimum expectations for IA among target audiences. In this vein, SPIA is now considering a follow-up activity oriented towards establishing a greater degree of consensus from investors as to their expectations regarding ex-post IA. For the purposes of eliciting client opinions of different

[1] Evenson, R. and D. Gollin (eds.). 2003. Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI; and Gardner, B. 2003. The CGIAR at 31: An Independent Meta-Evaluation of the Consultative Group on International Agricultural Research. Washington, DC, USA: World Bank, Operations Evaluation Department; Anderson, J.R. 1985. Summary of International Agricultural Research Centers: A Study of Achievements and Potential. Washington, DC: CGIAR Secretariat.

scenarios, studies and standards, a workshop is being planned with donors and impact assessment practitioners to supplement and build on a previous conference sponsored by SPIA and CIMMYT.[2]

SPIA is grateful for critical external input provided by a number of individuals during different stages of the study. Early discussions with Alex McCalla, Jim Ryan, and Bruce Gardner provided ideas for the study design and approach. Finalization of this report involved a three step process of review by SPIA Members; review by six external experts; and review and approval by the interim Science Council. In particular, SPIA appreciates the critique and valuable insights provided in the comments by Dana Dalrymple (USAID), Bruce Gardner (University of Maryland), Peter Matlon (Rockefeller Foundation), Mandi Rukuni (W.K. Kellogg Foundation), Meredith Soule (USAID) and Dunstan Spencer (independent consultant).

SPIA commends David Raitzer for a thorough, perceptive and innovative report produced on a limited budget. It is recognized that the results generated provide an excellent preliminary view and an opening for further, more detailed work on the returns to the overall investments in the CGIAR Centres. At the same time, this study provides the starting point for developing a set of standards and criteria for judging the plausibility and credibility of impact assessment related to agricultural research. As a result of this study's findings, SPIA concluded that there is need for expanded IA coverage in specific programme areas, such as natural resources management. In fact, such activity is presently part of the SPIA portfolio.

Hans Gregersen
Chair
CGIAR Science Council, Standing Panel on Impact Assessment

[2] See Watson, D.J. 2003. Why has impact assessment research not made more of a difference. Proceedings of an International Conference on Impacts of Agricultural Research and Development in San Jose, Costa Rica, 4-7 February, 2002. Mexico, DF: CIMMYT. The new proposed workshop would focus more on CGIAR donors and their needs. At present it is envisioned that participating donors would be asked to present short summaries of content, strengths, weaknesses, and points for improvement of specific CGIAR ex-post IA studies. This would help to define in clear and definitive terms the needs and expectations of some of the primary users of ex-post IA studies. Patterns of expectations evident in the conference would be distilled into minimum IA standards broadly acceptable to IA audiences. In turn, with these standards defined, it will be possible to revisit the wide spectrum of estimates developed in the present analysis, and 'zero in' on a more precise benefit range.

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centres of the CGIAR

David A. Raitzer

Prepared on Behalf of the CGIAR Standing Panel on Impact Assessment

SCIENCE COUNCIL SECRETARIAT

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
September 2003

TABLE OF CONTENTS

Figures

Tables

ACKNOWLEDGEMENTS

In the absence of key inputs from a number of individuals, the present report would have not been possible. First, I must express my sincere appreciation to Hans Gregersen, Chair of the CGIAR Standing Panel on Impact Assessment (SPIA), and Tim Kelley, SPIA Secretary, for critical guidance and feedback regarding the methodological approach pursued. I gratefully acknowledge the innumerable discussions of impact assessment methodologies held with these two individuals while conducting this analysis, which were crucial to development of criteria for the study review process. Thanks must also be expressed for the valuable suggestions and support provided by Shellemiah Keya, Executive Secretary of the CGIAR interim Science Council. In addition, I appreciate the assistance with formatting for final publication provided by Oksanna Hatalak and Tanya Alexander.

While reviewing individual studies, I received helpful input and clarifications from Derek Byerlee, Timothy Dalton, Robert Guei, Nancy Johnson, Aden Aw-Hassan, Victor Manyong, Phil Pardey and Greg Traxler. These responses greatly assisted my review efforts.

SPIA Members Ruben Echeverría and Herman Waibel, as well as interim Science Council Member Alain de Janvry offered useful suggestions regarding an early draft of this paper. Many thanks must also be expressed for the astute comments provided by the six external expert reviewers. These reflections have improved the cogency of this assessment, and have helped to contextualize the main findings.

EXECUTIVE SUMMARY

While the Consultative Group on International Agricultural Research (CGIAR) has been long considered a driving force behind the successes of the “Green Revolution,” no prior study has attempted to develop an aggregate estimate of the value of the System’s impacts. However, economic ex-post impact assessments (IAs) have been conducted for some of the most outstanding individual innovations of the System. This study synthesizes these specific benefits estimates, and sets such against total investments in the CGIAR, with the benefits from all other CGIAR activities omitted, so as to derive estimates for five different aggregate benefit-cost scenarios.

Economic impact studies for inclusion in the meta-analysis were selected based on a literature survey of publications databases, examination of reference lists from prior studies and scrutiny of International Agricultural Research Centre publications. Since impact assessment has been pursued in a largely decentralized manner, standards and approaches differ significantly among studies, and, hence, a critical review process was necessary for determining the reliability of generated results. To develop the conceptual grounding for the review process, best practices were identified for economic impact assessments.

Two overarching principles for evaluating study reliability- 1) transparency and 2) demonstration of causality, as well as accordant criteria and indicators, were developed from the identified best practices. Transparency was represented by three criteria: 1) clearly derived key assumptions, 2) comprehensive description of data sources, and 3) full explanation of data treatment. Demonstration of causality was represented by five criteria: 1) representative data set utilized, 2) appropriate disaggregation, 3) adequate consideration of mitigating factors, 4) plausible counterfactual developed, and 5) precise institutional attribution.

Using these criteria, five benefits scenarios were developed. These scenarios include 1) a scenario only including highly rated “significantly demonstrated” studies that empirically attribute benefits to specific activities of the CGIAR, rather than arbitrarily partitioning benefits from efforts in collaboration with partners, 2) a conservative scenario of only highly rated “significantly demonstrated” studies, 3) a selection of “plausible” studies meeting minimum standards for the criteria described above, 4) a “plausible, extrapolated to the present” scenario in which benefits for the crop genetic improvement studies were assumed to continue from the study period to the present (end of 2001) and 5) a “plausible, extrapolated through 2011,” which assumes that the products of current research will continue to be realized at present rates through 2011.

Against an aggregate investment of 7,120 million 1990 US dollars (6,900 million of investment in the CGIAR, plus relevant pre-CGIAR costs) from 1960 through 2001, all scenarios produced benefit-cost ratios in substantial excess of one, based on benefits accruing from 1972 – 2001. Including only “significantly demonstrated” studies that empirically attribute CGIAR derived contributions to collaborative efforts results in a ratio of 1.94, while if all “significantly demonstrated” studies are considered, with assumed attributive coefficients applied, the ratio rises to 3.77. The “plausible” scenario results in a ratio of 4.76, while when extrapolated to 2001 this rises to 9.00, and extrapolated through 2011, this becomes 17.26. Since costs are distributed over the benefit period, and many benefits peaked in the early 1990s, the discount rate applied only significantly affected generated ratios in the extrapolative scenarios.

The true value of benefits arising from the CGIAR is probably in excess of even the upper bounds of the results demonstrated here, as only a small subset of System impacts have been assessed. To illustrate this point, 98.1% of “significantly demonstrated” and 93.4% of “plausible” benefits were generated by just three research areas – cassava mealybug biocontrol, breeding of spring bread wheat and modern varieties of rice. Anecdotal evidence suggests that these are not the only areas of CGIAR research success, so there is substantial scope for expanded impact coverage, and better illustration of how System activities influence target beneficiaries. Furthermore, even where economically assessed there still remain significant opportunities for improving the methodological rigour, comprehensiveness and transparency of System assessments.

The diversity of methods employed among Centres and research programmes appears to indicate that additional resources for impact assessment leadership at the System level would offer considerable potential to improve consistency and raise analytical standards. However, for such leadership to be highly effective, it will be necessary for the “clients” of impact assessments to articulate expectations for substantiating different types of impact claims. In the absence of such, it is difficult to select one of the six scenarios as most “accurate,” and the “true” benefit-cost ratio of the CGIAR investment will remain unresolved.

1. INTRODUCTION

The Consultative Group on International Agricultural Research (CGIAR) was formed in 1971 to help foster technical solutions to production-related constraints affecting developing-country agriculture, through international applied research activities. Initially composed of four commodity-oriented International Agricultural Research Centres (IARCs), funded by three multilateral cosponsors, several non-governmental organizations and a smattering of bilateral donors, the visible successes of the System allowed it to quickly expand in size and scope of research agenda. Presently, the Group has 62 members supporting 16 IARCs, with research foci ranging from crop breeding to forest policy and strengthening of National Agricultural Research Systems (NARS).

Since establishment, the CGIAR community has invested approximately US $ 6.9[1] billion (2001 inclusive, in 1990 dollars) in various research and research related activities. According to Anderson (1998), the Group has been recently "threatened by a downturn in total real commitments that mismatches the increased demand arising from the recent expansion of the scope of the system." Despite widely acknowledged successes, funding for the CGIAR has stagnated in real from 1988 through 2001. Such a trend may imply that economic impact assessments conducted for the System thus far have not adequately convinced donor audiences that the System portfolio is an exceptionally efficient and productive investment. Accordingly, one principal point of scepticism expressed by donors regarding previous evidence of efficacy is that such have "cherry picked" research successes, while ignoring the costs of failures (TAC Secretariat, 2001). The present analysis offers an answer to such criticism by compiling highly reliable estimates of widely-recognized benefits, and setting such against *total* investments in the System to-date.

The question addressed in this study is: Do the documented benefits from CGIAR research justify the total investment in the CGIAR so far? Thus, the basic objective is to derive a set of plausible and highly credible aggregate estimates of the benefits accruing from System innovations, and to set such against the present value of the entire CGIAR expenditure. In so doing, all undocumented benefits or documented benefits that do not meet certain criteria for selection make no contribution to the numerator of aggregate benefit-cost ratios, while all System costs (including those of facilities and related activities, such as communications or training) are included in the denominator. Consequently, the present approach is biased towards conservatism by the fact that impact assessment has only been applied to a small proportion of CGIAR research activities. Ratios derived through this method should provide meaningful insights into the aggregate effectiveness and efficiency of the CGIAR investment. Most importantly, these estimates avoid the common criticism that only successful projects are often compared, while the costs of the unsuccessful are ignored.

[1] Additional investments are made in the first four IARCs prior to the CGIAR's establishment, and those relevant to the present analysis are included in cost figures cited later in this document, which appear as 7,120 million 1990 US Dollars.

1.1 Approach

The overall approach used is presented in Figure 1. This study begins by providing background on methods typically applied for assessing the impacts of agricultural research, particularly in terms of basic conceptual approaches behind economic impact assessments. The methodology applied in the present study is subsequently described, beginning with the means by which studies were identified. Next, a model of "best practices" for economic impact assessments (IAs) is developed, so as to have a consistent basis for reviewing the diverse set of impact assessments identified. These "best practices" are then used a basis for developing specific criteria and indicators for assessing study quality. After the development of these criteria, the aggregation process for benefit values produced in individual studies, as well as the limitations of the approach taken in the present study are presented as a final step in the methodology section. Subsequently, the results section describes key outcomes, including benefit-cost ratios under six scenarios. Then, the discussion section examines the significance and likely accuracy of the benefit-cost ratios produced under the different scenarios. In turn, the discussion focuses on the distribution of benefits generated and the pathways by which such diffused. Next, key aspects of and trends in the quality of the reviewed studies are discussed. Finally, the conclusions section draws upon key points of the discussion to recommend future actions for the CGIAR and its Standing Panel on Impact Assessment (SPIA).

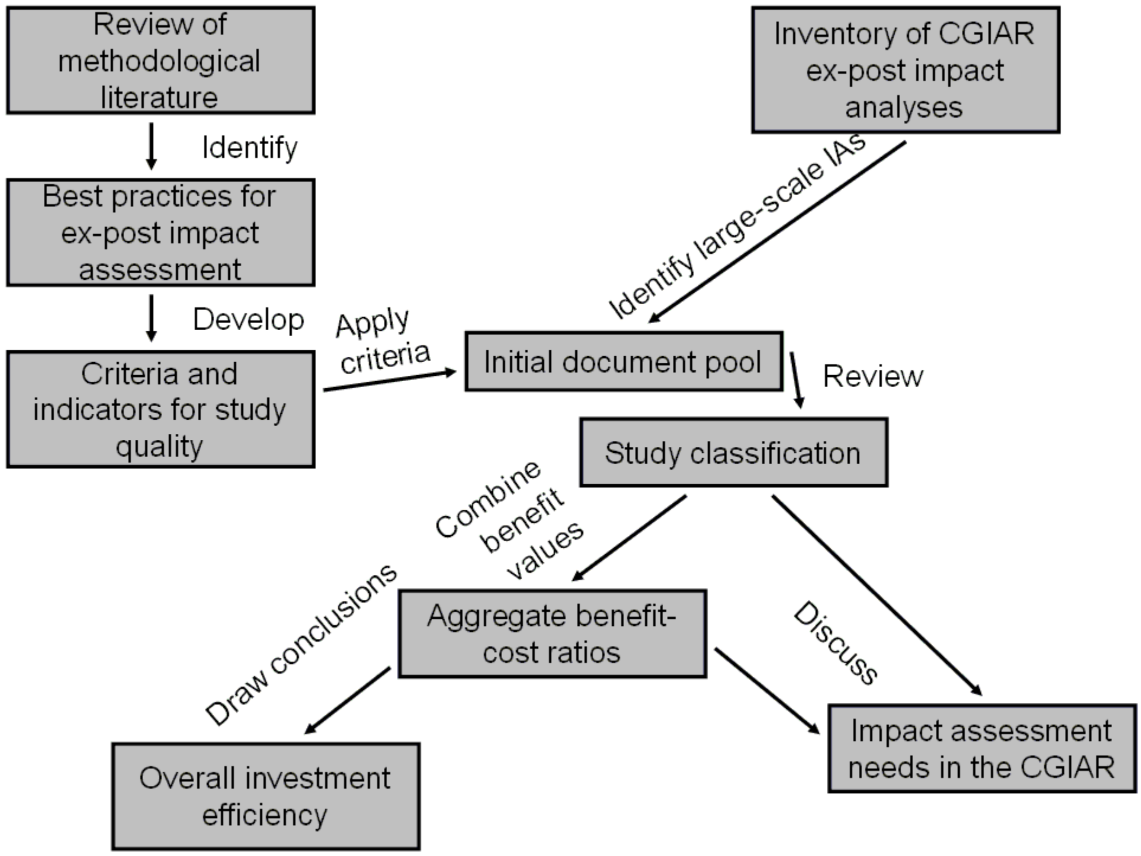


Figure 1. Stylized analytical and methodological pathway of the benefit-cost meta-analysis of the Consultative Group on International Agricultural Research

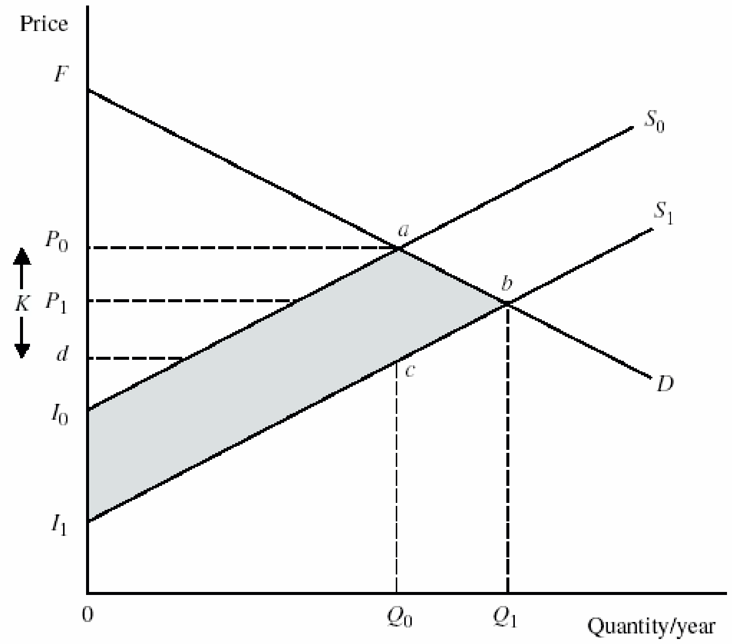
1.2 Background on valuing the benefits of agricultural research

To provide some background for the methodology applied in the present meta-analysis, it is illustrative to offer an overview of the general economic impact estimation techniques applied to agricultural research. In order to delve into impact estimation methodologies, it is first helpful to explain the generic impact pathway model conceived for productivity-enhancing research. Schultz (1964) was one of the primary economists to first argue that subsistence agricultural production systems in the developing world are technically efficient, and that farmers maximize profit given available technology. This is a principal premise behind the benefits envisioned as a result of agricultural research, as if producers are efficient optimizers, only by raising the production frontier through the transfer of new technology will productivity be improved. Research thus offers a key means to develop new technologies which will raise this productivity frontier, so that higher quantities of agricultural products will be supplied at any given price, and that this enhanced supply will drive down prices. Consequently, this improved producer income through higher productivity, coupled with lower prices, which raise consumer purchasing power, will underpin economy-wide growth. Agricultural development as the engine underpinning broad economic progress has now been accepted as a central tenet of modern development theory, as indicated by Shultz's reception of the Nobel Prize in 1979 (for these and subsequent ideas).

Generally, economic impact, as commonly assessed for agricultural research at large spatial scales, is not on poverty, but is a quantification of gains resulting from productivity improvement. It is often taken as granted that such productivity increases will foster gains throughout the broader target economies, and thereby achieve ultimate goals of poverty alleviation. Thus, the values presented are, in most cases, quantifying intermediate, rather than mission-level impacts. In this sense, impact assessment for agricultural research differs from the World Bank's definition of IA as "intended to determine more broadly whether the program had the desired effects on individuals, households and institutions and whether those effects are attributable to the program intervention" (Baker, 2000). This is not to say, however, that this mission-level focus has not been investigated in impact assessments conducted within the System, as numerous studies have explored impacts of more immediate proximity to poverty than gross research benefits (David and Otsuka, 1994; Hazell and Haddad, 2001; Hazell and Ramasamy, 1991; Kerr and Kolavalli, 1999; Renkow, 1993). However, these are largely case studies, and there are few well developed methods by which such impacts may be scaled up to large geographic areas or represented by singular aggregate statistics.

Within the CGIAR, there has been no strong formal institutional control mechanism for impact assessment at the System level. Integration and guidance at the System level were initiated with the inauguration of the Impact Assessment and Evaluation Group (presently entitled the Standing Panel on Impact Assessment, or SPIA) in 1995, prior to which the individual IARCs developed their own IA standards, approaches and techniques relatively autonomously to deal with their own research. Thus, there is great diversity in standards and techniques applied in the Centres.

Most economic impact assessments of agricultural research rely on economic surplus techniques, which in some cases are fed by the results of econometric methods. Economic surplus techniques build upon the approach first utilized by Griliches (1957) in his pioneering study on hybrid corn, in which adoption of a technological innovation fosters a downward shift in the supply curve, through reductions in the unit cost of production. Costs per unit of production may be lessened through loss reductions or increased yield potential. The value of this per hectare increase in productivity attributable to the innovation, multiplied by the adopting area planted during a single year, then gives the gross annual research benefits. Figure 2 illustrates the market effects of increased productivity for a simple "closed economy." The counterfactual scenario is embodied in continued application of the previous, or next-best technology, represented as supply curve S_0, while the curve with the innovation applied is S_1. Benefits may be presented in an aggregate social form, including changes to producer as well as consumer surplus[2], or if price elasticities of demand and supply are utilized, may be partitioned between producer and consumer groups. Benefits to the latter are usually determined to be prevalent through reductions in food prices for "closed economy" models, represented here as $Q_0*\Delta P + (\Delta Q*\Delta P)/2$.



Source: Alston and Pardey 1996.

Figure 2. Representation of a general model for valuing technological innovation within a "closed economy"

Open-economy models are also often used, and these do not base benefits on consumer price reduction, but assume that supply does not affect average prices, due to the presence of imports. Under such assumptions, benefits are measured through the value of increased production or inputs saved per unit of production, and are implicitly assumed to be received by producers.

Alternatively, econometric techniques allow for arguably more precise estimation of research benefits through explicit specification of endogeneity versus exogeneity of included variables during statistical analysis. However, econometric techniques are also arguably more

[2] If elasticities are not utilised, annual prices or an average price for the period of benefit estimates may be often used to encompass the price effects of supply shifts attributable to research, as total benefits are not necessarily sensitive to elasticities of supply and demand (Alston *et al.*, 1996).

liable to the assumptions underlying the model of relationships utilized, as spurious correlations may be interpreted to imply causality. In addition, the step of valuing additional output or savings in inputs essentially still relies upon an implicit economic surplus analysis (Alston *et al.*, 1996). In this regard, this approach is a derivative of the economic surplus technique, rather than a separate set of methods, as some authors may contend.

The scope of economic impact assessments is usually that of a single innovation or series of innovations. Since research is an uncertain process characterized by many "dry holes" producing little beneficial impact, and a few "gushing wells" producing substantial benefits, analysis at the project or programme level cannot necessarily provide meaningful insights at the System level, unless the costs of associated unproductive investments are considered, as well. Accordingly, the CGIAR's Task Force on Impact Assessment recommended that there was need for impact assessment at the System level as well as at the Centre level, to ensure that the costs of both successful and unsuccessful projects within the CGIAR are included in assessing research impact (Özgediz, 1995).[3] The present study should help to operationalize this recommendation.

2. METHODOLOGY

2.1 Literature inventory and selection of studies for inclusion

The present meta-analysis utilized a critical literature review of ex-post economic impact assessments to obtain benefit values for aggregation in the numerator of the CGIAR benefit-cost ratio. Studies were sourced from recognized peer-reviewed books and journals, as well as publications directly produced by individual IARCs. The initial selection of publications for review was based upon a comprehensive inventory of impact assessment literature, through searches in relevant literature databases, such as Agricola, Agris and EconLit. In addition, the publications lists of each of the CGIAR Centres was reviewed, and citations lists from relevant literature were perused, in order to obtain as comprehensive a selection of benefit estimates as possible.

During this search process, global or macro-regional benefits studies were more heavily sought than smaller-scale studies, and a minimum cumulative ex-post benefit estimate of $50 million was required for initial inclusion. The minimum cut-off value was a product of the significant value of the total CGIAR investment to date, as benefit estimates below $50 million (0.7% of the funds invested) have little impact upon the aggregate benefit-cost ratio. Although necessary for the efficiency of the review process, it should be noted that this criterion alone was heavily restrictive, and excluded the vast majority of published IAs.

Only studies published after 1989 were included in the initial document pool, as lag periods between impacts and data collection (often 3-4 years), as well as between research activities and impacts (commonly more than a decade), mean that studies from prior to this year encompass little time for the effects of CGIAR activities to have become evident, or that such studies are likely to include significant effects from pre-CGIAR research investments. While this temporal criterion may have resulted in the exclusion of some significant

[3] The term "unsuccessful" is used in the economic sense related to intended ultimate beneficiaries. It is of course true that such "unsuccessful" research can be highly successful from the point of view of expanding the general scope of scientific knowledge, which in turn may provide the necessary foundation for future discoveries.

documented impacts, such restriction is necessary to ensure that estimates are conservative, and do not mostly reflect the prior contributions of other research institutions.

To restrict benefits broadly to target regions, only analyses covering lower and middle income countries were included. CGIAR participation or connection to the analytical process was the final criterion for inclusion in the initial study pool, as the System's institutional contribution to impacts cannot be reliably determined without obtaining programmatic information from the IARCs.

2.2 Overview of studies in initial document pool

Most of the studies in the initial selection have been independently produced by individual Centres. Thus, there is significant diversity in methods employed, and key assumptions differ significantly among the studies (Cooksy, 1997). Since the purpose of the present study is essentially to derive a **credible minimum** aggregate benefit level for the CGIAR investment, it is essential that a high degree of confidence may be placed in included values. To do so requires "filtering" for only the most reliable studies. As a result, a critical review process was utilized to disaggregate by characteristics of the methodologies employed after the initial pool of documents had been selected (Figure 3).

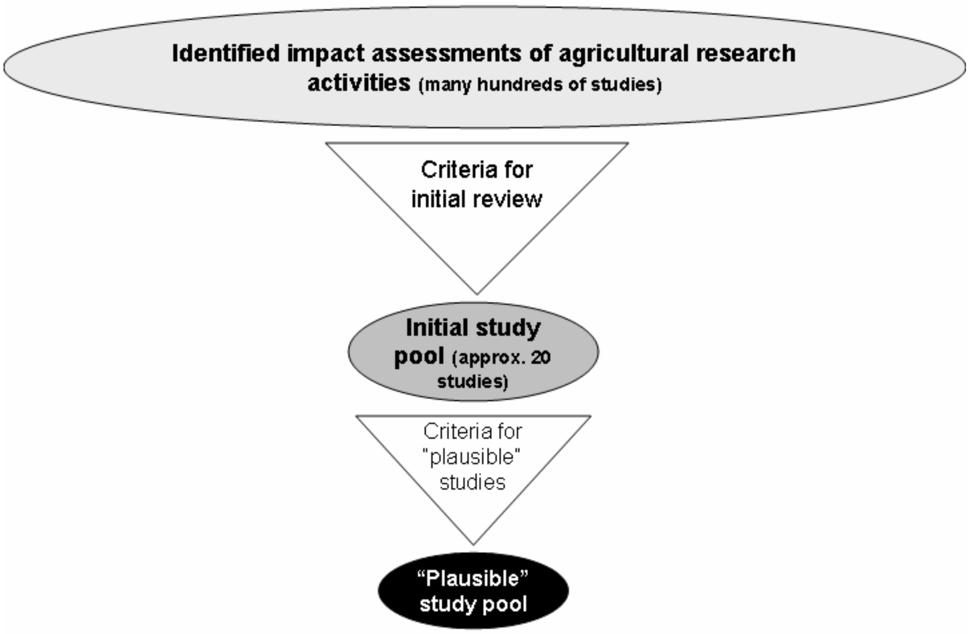


Figure 3. Diagram depicting the overall process used in the present meta-analysis to select for "plausible" economic impact studies

2.3 Data collection for review process

Once an initial literature selection was made, key aspects of the methodologies of each of the studies were recorded in an Access database. This database then allowed for qualitative assessment of the methodologies employed against criteria enumerated below. A list of database fields is included in Appendix I.

2.4 Description of "best practices" in economic impact assessment

In order to develop criteria for assessing and including reviewed impact assessments, a general impact assessment model of "best practices" has been developed in the present study through literature review. Although it is recognized that research effects are diverse in nature and that no single linear model can be extrapolated to analyse all possible means of impact, the transparency of the review process is enhanced through specification of a set of standards from which review criteria were drawn. Furthermore, while it is recognized that no real-world study could completely satisfy all of the described conditions, this model of best practices provides the conceptual grounding for categorizing the studies to be included in the present analysis. Even though this model is technically an intermediate result, it is included here, as it comprises a necessary step for the subsequent review of included studies.

For the purposes of the present analysis, ex-post impact assessments can be characterized as typically composed of five interlinked subcomponents:

1. Description of the focal institution's (the institution for which impact is being assessed) participation in innovation development and the impact pathways of the research process.

2. Empirical estimates of observed trends in technological adoption and productivity.

3. Attribution of observed productivity trends to different relevant causal factors.

4. Development of a hypothetical counterfactual scenario, which supposes the course of action without development and adoption of the innovation (when contrasted with observed productivity trends with the innovation, this forms the basis of impact claims).

5. Economic valuation of the benefits attributed to the research process.

2.4.1 Impact pathway elucidation

Impact pathways may be defined as the conceptualized connections between an intervention and effects on the broader social and physical environment. When impact pathways are defined and described in ex-post IA, a theoretical framework is set for the broader analysis. The research process upon which the IA focuses and the IARC or System role in this process should be delineated in detail, along with relevant innovations produced. Thus, the development of the analysed innovation should be described precisely, so that the proposed impact pathway can be reliably credited, and contributors attributed. It is particularly important that the intended advantages of the developed product be presented, as such will essentially establish conceptual boundaries for the assessed impact pathways. According to Anderson (1997) "to link…elements of agricultural development…to enhanced productivity and welfare…in a consistent framework for measuring productivity is task enough" in and of itself!

2.4.2 Estimation of adoption and productivity trends

Productivity effects can often be collapsed into two factors – adoption and increased efficiency. Both aspects require significant amounts of field data to derive meaningful results. In the absence of a comprehensive and representative empirical basis for claimed impact levels, benefit levels estimated, while potentially plausible, have little true meaning, as realistic assumptions for one agro-environment may not be necessarily extrapolated across agro-ecological conditions. Hence, it is essential that representative sampling procedures be utilized, which are adequately inclusive of the range of conditions and resource endowments under which the innovation is employed. As Maredia *et al.* (2000) note in their *Tour of Good Practice* "estimates of research benefits should be disaggregated by commodities, production environment, or geographical basis if the parameter estimates are different for different components of a research programme." The sample size should be sufficient for deriving reasonably (given time and resource constraints) precise and repeatable figures, and the population should be randomly sampled, to prevent bias.

The data gathering process should ideally encompass the concept of multiple-source-verification or triangulation, which is defined by Baker (2000) as a set of "procedures that permit two or more independent estimates to be made for key variables." In other words, the results of one data collecting process may be validated through the application of other methodologies, which can produce the same kinds of data. Similarly, Maredia *et al.* (2000) recommend that impact assessments "combine technical, scientific and economic information from a number of sources." For example, adoption data derived from seed sales estimates may be validated through field surveys, and yield information derived from experiment station data should be corroborated with surveys or on-farm trials, as even relative yield gains may be greater in the breeding environment than in farmers' fields.

2.4.3 Attribution of causal factors

It is critical that an impact assessment attempt to demonstrate causality, rather than proceed on the basis of assumed relationships. Correlations do not establish causation when presented in isolation, so the data gathering process must illustrate that the target innovation is a primary causal factor. To do so requires that mitigating influences be identified, and the role of such in affecting productivity changes be methodically assessed. For example, often adoption of a technological innovation acts as a catalyst for adopting other related technologies, by making investment in complementary inputs more profitable (e.g. semi-dwarf varieties eliminate lodging, thereby allowing higher rates of fertilizer application). Thus, the respective roles of complementary contributions to yield increases, should be assessed empirically, rather than arbitrarily supposed.

Furthermore, it should be recognized that many factors interact to influence productivity trends, and that these factors may be exogenous to the agricultural sector. Macroeconomic and trade policies, infrastructure, human capital development and property rights regimes may all catalyse changes in agricultural output and performance, since farmers adaptively optimize production techniques according to external conditions. Given this context, it is a crucial challenge for impact assessment to isolate changes in the productivity frontier from changes in factor use within the same technological boundaries. In so doing, it is particularly important to accurately represent how shifts in the boundaries of the productivity frontier transpire, as inaccurate representations of these shifts have seriously skewed results in the past (Nin *et al.*, 2003).

2.4.4 Counterfactual[4] development

According to the World Bank's *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners* "to ensure methodological rigor, an impact evaluation must estimate the counterfactual, that is, what would have happened had the project never taken place or what otherwise would have been true ... [as] determining the counterfactual is at the core of evaluation design" (Baker, 2000). In many cases, even in the absence of CGIAR activity, it is likely that the NARS would still be producing research products, which would be generating a certain degree of impact. Thus the additionality of the CGIAR ought to be systematically estimated, so as to avoid crediting the CGIAR System with NARS products or vice-versa. This should be based on empirical assessment, and econometric methods offer one possible means to do so. Even in the absence of NARS research efforts, it is likely that yield gains would have been achieved through private sector research, farmer innovation, changes in crop management or factor substitution, and the counterfactual should include these elements. Ideally, such should be based on empirical analysis of farmer substitution opportunities and patterns, and should be illustrated as the "next best" option for maximising utility. The counterfactual, though often only implicitly developed, is equal in importance to the derivation of observed productivity gains for determining the accuracy and precision of results, as the difference between the productivity estimates under the two scenarios constitutes the basis of claimed benefits. Since this subcomponent is of such high relevance for the accuracy of benefits assessed, the methodology for counterfactual derivation should be presented in extensive detail.

Preferably, the counterfactual should be based on empirical analysis of comparisons between a "control group" and roughly equivalent populations of beneficiaries (Baker, 2000). Through socio-economic mapping, in combination with maps of ago-ecological conditions and factor-endowments, it should be possible to isolate groups of non-adopters and adopters with similar characteristics, and trace productive profitability over time. In addition, it should be possible to trace marketing channels to analyse supply changes induced by new technologies, and the effects of these for consumers.

2.4.5 Economic valuation

Once productivity impacts have been estimated as the difference between the observed adopter estimates and the counterfactual, they should be assigned an economic value. Establishing a proper price for the change in production is often impeded by market distortions, such as monopoly buyers, competition with subsidized exports, export taxes, or by poor infrastructure. "Economic prices that are appropriately adjusted to reflect policy distortions in the output market" should be utilized for valuing productivity changes, according to Maredia *et al.* (2000). The most basic price for utilization is the world market price, and this is an effective means of valuing export commodities, since price distortions are most likely to depress, rather than inflate international market prices.

[4] The counterfactual is the hypothetical course of events that would have taken place in absence of the assessed activity or contribution.

However, crops primarily cultivated for subsistence use are much more difficult to value, especially when there are few possibilities for substitution. The most representative price for such commodities is probably the domestic price. To maintain credibility, it is best to be conservative in estimating prices, so as to not overvalue productivity increases. In addition, the price effects of supply changes should be assessed through empirically derived demand and supply elasticities. Such are necessary for the development of meaningful analyses, which allow for the distribution of benefits to be assessed. Furthermore, if secondary research impacts are to be estimated, such as reductions in deforestation, it is necessary that the price effects of supply changes be calculated. It is also potentially valuable to spatially disaggregate price effects of supply changes, as marginal rural areas served by poor infrastructure may experience different levels of price change than do major port cities, and such has substantial implications for the distribution of benefits.

Furthermore, when changes in the use of factors with significantly distorted market prices are analysed, adjustments should be made so that the social costs of alterations in input use are included in benefit assessments, through the application of shadow prices. For example, changes in pesticide levels applied to cultivars should be valued at the pesticide price plus incurred additional external costs related to human health or the natural environment. Similarly, Alston *et al.* (1996) recommend that "total benefits [when accompanied by externalities] are given by deducting the amount of the increased external cost from producer benefits." Once all cost factors are converted to shadow prices reflective of social values, complementary effects may be considered in economic terms.

2.5 Criteria for study selection

After "best practices" in impact assessment have been described, it was possible to develop a systematic qualitative review process to assess the degree to which such practices have been realized in specific studies. As mentioned previously, the reviewed studies vary greatly in methodological transparency and sophistication. Consequently, since the reviewed studies are of such variable standard, this process was needed to determine the level of confidence and conservativeness with which results may be used.

To develop a structure within which the studies could be reviewed, a hierarchal framework from principles to criteria to indicators was developed (Table 1). In certain cases, criteria and indicators were merged into a single metric. The two overarching principles for the review of assessments were 1. transparency and 2. demonstration of causality, while the former of which was a necessary condition for the later.

2.5.1 Transparency: criteria and indicators

Since the ability to understand the basis of derived results is a requisite condition for placing confidence in findings, it is imperative that credible studies are characterized by transparency. For the purposes of this study, transparency as a principle was represented by three broad criteria (Figure 4):

1. clearly derived and explained key assumptions
2. comprehensive description of data sources
3. full explanation of data treatment.

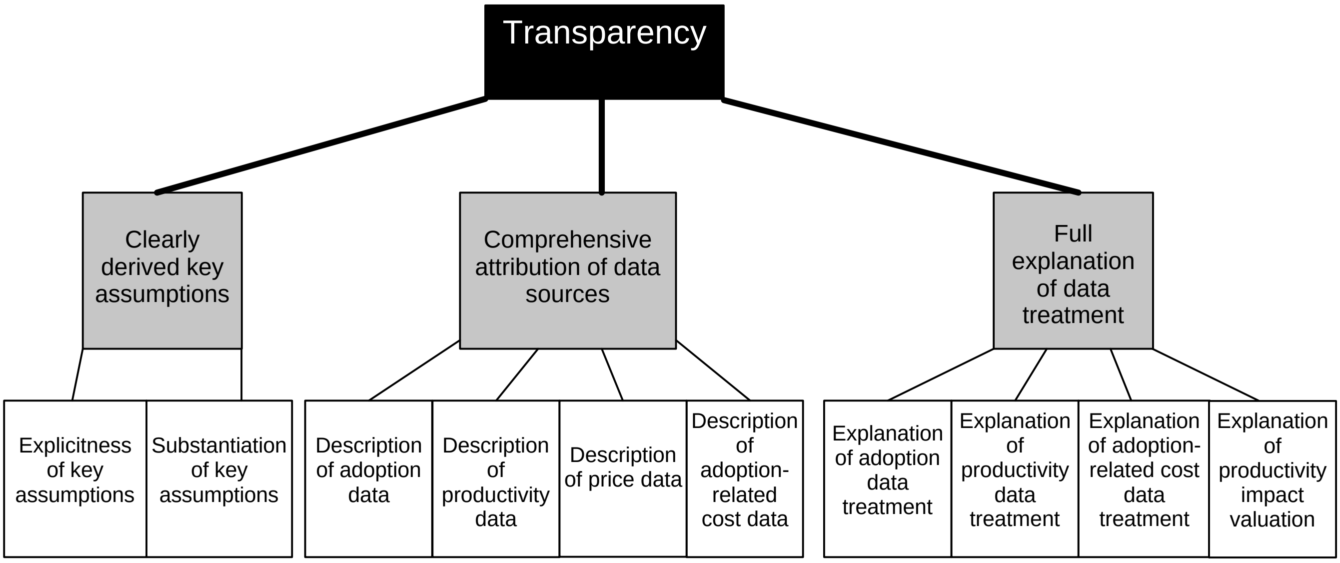


Figure 4. Hierarchical relationship between criteria and indicators for assessing the transparency of reviewed studies

Clearly defined key assumptions

This criterion was represented by two qualitatively assessed indicators – explicitness of key assumptions and substantiation of key assumptions. The explicitness indicator refers to how openly the study defines which aspects of the analysis have been derived from expert opinion or presumption by the author(s). Substantiation refers to whether a logical basis or citation has been provided to authenticate these untested assumptions.

Comprehensive description of data sources

Under this criterion, four indicators were enumerated, including description of data sources for extent of adoption (when relevant), productivity effects, costs associated with adoption and prices for valuing productivity changes. For each of these factors, it was noted whether all apparent sources of data were specifically cited. General or vague references, such as “productivity data were cited from studies,” with no in-text citations listed, or “interviews were conducted,” with no sample size presented, were rated poorly, while precise citations for all specific data utilized were rated highly.

Full explanation of data treatment

Four indicators were derived for this criterion, in a similar manner to those of the attribution criterion. These include explanations of adoption, productivity, costs and valuation (including discounting/deflation). The ideal against which studies were evaluated was the provision of sufficient information to allow repetition of the methodology used for processing each of these kinds of information.

2.5.2 Demonstration of causality: criteria and indicators

For a study to make a credible claim of impact, it is essential that a causal linkage be established between the intervention and claimed effects. To address the degree to which the reviewed study demonstrated causality, five criteria were developed:

1. Representative data set utilized

2. Appropriate disaggregation
3. Adequate consideration of mitigating factors
4. Plausible counterfactual scenario developed
5. Precise institutional attribution.

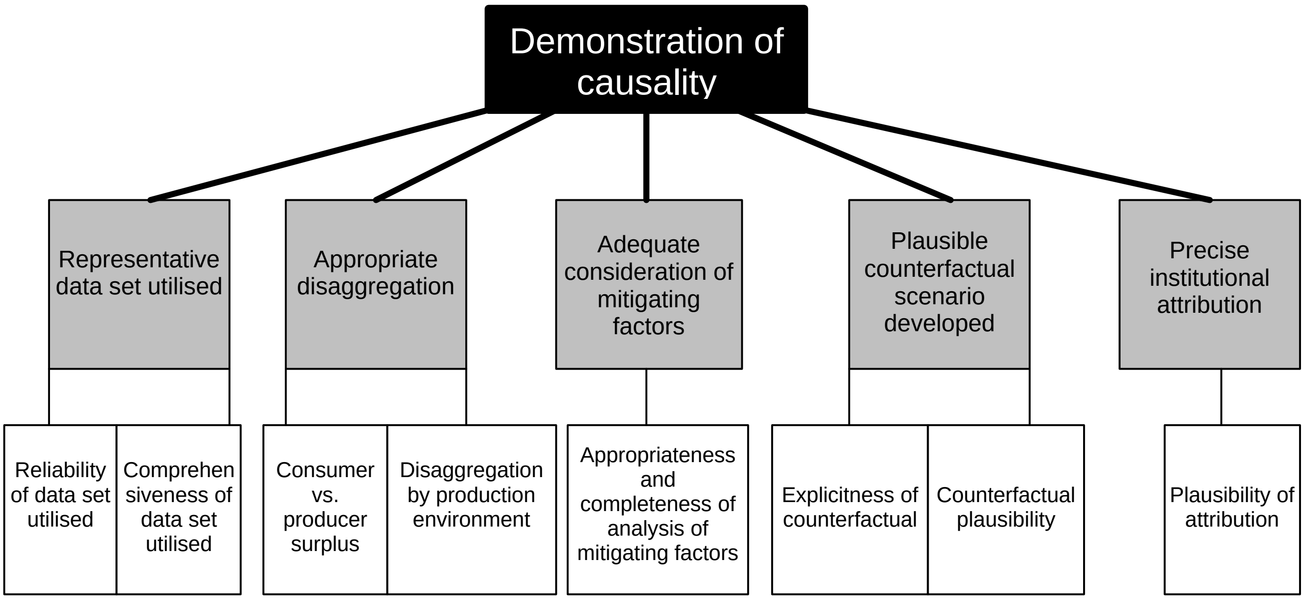


Figure 5. Hierarchical relationship between criteria and indicators for assessing the demonstration of causality within reviewed studies

Representative data set utilized

This criterion refers to whether the data set utilized for generating adoption and productivity estimates was likely to accurately and precisely represent target populations, and was represented by two indicators – reliability of data set utilized and comprehensiveness of data set utilized. Reliability refers to whether the methods applied for generating observations were likely to encompass significant bias or accurately represent analysed trends. Triangulation, with multiple-source validation applied was most highly regarded, while expert opinion as a sole basis was given the lowest score. Under comprehensiveness of data set utilized a rating was given for the sample size used, both in terms of geographic and temporal coverage, particularly with respect to the basis of claimed productivity effects.

Appropriate disaggregation

To assess the degree to which impact heterogeneity is considered in the reviewed studies, two indicators were used for evaluating the fulfilment of this criterion – disaggregation by agro-environment and disaggregation of surplus generated (between and among producer and consumer groups).

Adequate consideration of mitigating factors

Numerous causal factors apart from CGIAR derived research outputs may explain observed trends in productivity. For the purposes of this review, a single indicator denotes whether major relevant classes of mitigating influences (such as infrastructure, policy and crop management) were incorporated. The degree to which the relative contribution of these other factors was contemplated in estimating impact causality has been assessed for each of the studies.

Plausible counterfactual scenario developed

Two indicators were used to assess the counterfactual scenarios developed in the reviewed studies – counterfactual plausibility and explicitness of counterfactual. Plausibility of implicit or explicit counterfactuals indicates the degree to which the assumed course of events in absence of the innovation represents a realistic “next best” course of action. The relative explicitness of the counterfactual was used as a proxy indicator for the precision with which the counterfactual has been derived in the reviewed analyses. In utilising this factor as an indicator, it was assumed that explicit counterfactuals can inherently give a more precise representation of the changes that would be likely to occur in the “without” scenario, than can implicit counterfactuals based on technological contributions to changes between “before” and “after” scenarios.

Precise institutional attribution

This criterion was represented by a single indicator, which attempted to capture the plausibility of the attributive basis for crediting the involved IARC, for those studies that attempt to do so. When a study did not attempt to attribute the relevant Centre contribution to research products, it received the lowest score. Such should not be interpreted as indicating that this factor renders these studies of low reliability, but it does reduce the reliability of any IARC attributed values derived from these analyses, since the present analysis applied assumed attributive coefficients to derive the CGIAR portion of benefits estimated in such studies.

Table 1. Principles, criteria, indicators and rating examples for evaluating benefit-cost studies

Principle	Criteria	Indicator	Low rating	High rating
Transparency	apparent basis of key assumptions	explicitness of key assumptions	major assumptions underlying analysis are not defined	all major assumptions explicitly stated
		substantiation of key assumptions	explicit assumptions have no clear basis	explicit assumptions have logical justification and/or citation
	complete citation of data sources	citation of adoption data	unclear basis of adoption estimates	adoption estimates cited and/or methodology described
		citation of productivity data	unclear basis for productivity claims	productivity claims based on cited references or clear methods
		citation of adoption-related costs data	unclear empirical basis for deriving costs associated with adoption	estimates of adoption-related costs (where considered) cited or given logical justification
		citation of price sources	unexplained basis of commodity prices	cited basis for commodity prices
	full explanation of data treatment	explanation of scaling-up adoption estimates	no basis provided for adoption estimates	gathering process for adoption estimates defined
		explanation of scaling-up productivity estimates	unclear extrapolation from limited productivity impact data	clearly defined methodology for scaling-up site-specific impact estimates
		explanation of scaling-up adoption-related costs	unclear manner of incorporation of costs associated with adoption	costs considered (or not considered) in an explicit manner
		explanation of economic valuation	commodity prices used, discounting and deflating unclear	commodity prices used, discounting and deflating clearly presented
Demonstration of causality	representative data set utilized	reliability of data utilized	data sourced from uncorroborated expert opinion	data sourced from empirical studies or methods and validated through triangulation
		comprehensiveness of data set utilized	data sourced from a small set of unrepresentative sites	large number of sample sites representing the range of relevant agro-environments.
	adequate consideration of mitigating influences	appropriateness and completeness of analysis of mitigating factors	no mitigating factors considered	comprehensive consideration of all major relevant alternative causal factors
	appropriate disaggregation	disaggregation by production environment	only "average" conditions considered	heterogeneity in impacts appropriately captured
		consumer vs. producer surplus	gross benefits presented without analysis of surplus recipients	impacts disaggregated among different producer and consumer groups
	plausible counterfactual developed	explicitness of counterfactual	no "without scenario" presented	"without scenario" comprehensively developed
		counterfactual plausibility	counterfactual represents unrealistic, overly cynical, course of action	counterfactual represents realistic, likely and substantiated path of events
	precise institutional attribution	plausibility of attribution	no attribution attempted	empirically-based attribution derived through counterfactual

2.5.3 Rating of studies and aggregation scenarios

For each of these indicators, qualitative review produced ratings, which were aggregated and averaged for the two principles to derive a transparency rating and a demonstration rating. In the present study, two main levels of conservativeness were applied, and these are termed “significantly demonstrated” and “plausible,” respectively. The primary difference between the two main standards concerns the degree to which impacts must be proven before being quantified, as opposed to being assigned a reasonable value based on limited evidence. From these two main standards, one additional derivative of the "significantly demonstrated" scenario and two additional derivatives of the "plausible" scenario were drawn, for a total of five scenarios (Table 2; Figure 6).

Table 2. Characteristics of five scenarios under which benefits were assessed in the present analysis

	1. Significantly demonstrated & empirically attributed	2. Significantly demonstrated	3. Plausible (no extrapolation)	4. Plausible, extrapolated to the present	5. Plausible, extrapolated through 2011
Transparency	substantial	substantial	substantial	substantial	substantial
Causality illustrated	substantial	substantial	Limited	limited	limited
Attribution	empirical	empirical & assumed	empirical & assumed	empirical & assumed	empirical & assumed
Assumed attributive coefficients	NA	0.5	0.5	0.5	0.5
End period of benefits	based on study, < 2002	based on study, < 2002	based on study, < 2002	extrapolated to 2001 from final year of study	extrapolated through 2011 from final year of study

1. Scenario of "empirically attributed and significantly demonstrated" studies

Most of the reviewed studies did not take the step of attributing research products to different institutions via empirical means, as such has not been a commonly recommended practice in impact assessment (Maredia *et al.,* 2000). In lieu of such empirical attributive criteria, plausible assumptions were required to define the portion of collaborative research benefits that resulted solely from CGIAR activities. However, assumed attributive adjustments are speculative, and reduce confidence in results when only the isolated activities of CGIAR Centres are to be assessed. Thus, to be highly conservative, only those "significantly demonstrated" studies that take the extra step of empirically attributing the IARCs should be included in aggregate estimates. Since this scenario is also predicted on classification as "significantly demonstrated," this highly selective standard only incorporates studies that are both rated highly for transparency and rigour, as well as inclusive of an empirical basis for partitioning institutional credit.

2. Scenario of "significantly demonstrated studies"

The "significantly demonstrated" scenario is conservative, and requires that substantial evidence supports impact claims before they are included in aggregate figures. Requisite conditions for this classification include high ratings for transparency and impact demonstration. Fulfilling the criteria for the "plausible" scenario is a necessary but insufficient condition for the "significantly demonstrated" scenario.

3. Scenario of "plausible" studies

In this analysis, the "plausible" standard could be conceptualized as derived from an *a priori* position that CGIAR impacts are substantial, when such is supported by expert opinion, and that the role of IA primarily concerns quantifying the value of these effects. Meeting a substantive rating for transparency and a limited rating for demonstration qualifies studies as "plausible."

4. Scenario of "plausible" studies extrapolated to the present

Since the plausible scenario contains several single year benefit estimates for very large impacts of breeding research, which are truncated to the terminal year of the study period, benefit levels for these research areas may reasonably be expected to continue after the analysed year(s). Significant empirical evidence supports the contention that the benefits of varietal research have not dropped off significantly since these estimates were made, so this could be regarded as a plausible assumption (Reynolds *et al.*, 1999; Sayre *et al.*, 1997). For these scenarios, benefits were estimated to continue at the real rate of the final estimate year through 2001. In cases where such extrapolation could potentially cause benefits to be double-counted, such as when several sequential studies for the same research product are utilized, the more highly-rated estimate was used to cover the period.

5. Scenario of "plausible" studies extrapolated through 2011

The lag periods between investment and the realization of agricultural research impact mean that investment taking place at present will not often begin to have substantive social/economic effects until at least a decade from now. Furthermore, agricultural research products have a useful lifespan that may often exceed ten years. This implies that many of the productive effects of recent investments will not be yet realized, and that future benefits should be estimated to fully capture the effects of investments to-date. If it is assumed that research outputs from such investments are at least sufficient to maintain the level of benefits estimated by the reviewed studies for the next decade, such benefits may be plausibly extrapolated through 2011. Thus, in what may be regarded as the most optimistic scenario, the plausible real benefits from those research products which have been characterized by stable or rising benefit values in the late 1990s were carried forward at a constant level (prior to discounting) from the final study year through 2011. While this scenario helps to offer a fuller estimate of benefits than do the other four, the results are highly speculative, and are subject to a very high level of potential error, as previously unassessed research products are implicitly included. Thus, this is not truly an ex-post scenario, although it only attempts to include the effects of past research activities.

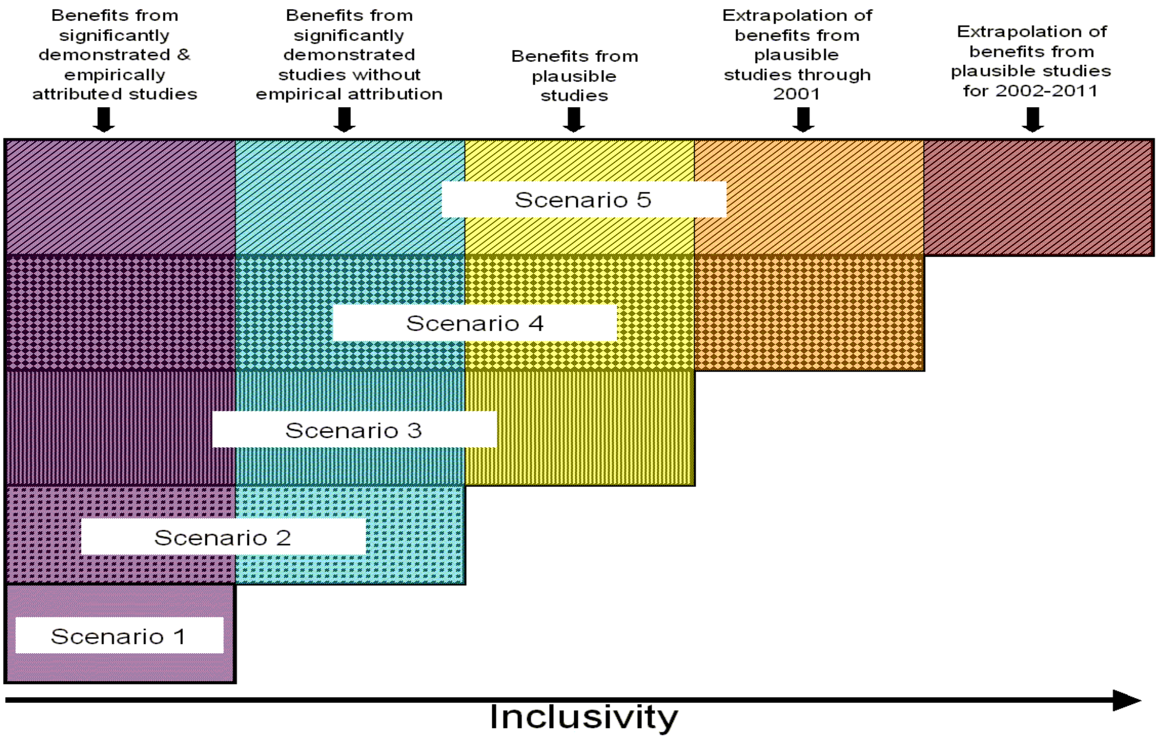


Figure 6. Additive inclusion of benefits in the five scenarios of the benefit-cost meta-analysis

2.6 Aggregation of results

2.6.1 Attribution between NARS and CGIAR

In many of the included impact studies, no attempt has been made to partition the benefits generated between NARS and IARC efforts. This presents a significant hurdle for the reliable estimation of efforts attributable to the CGIAR investment alone, as an explicit, detailed and substantiated counterfactual is necessary to precisely discern relative contributions. However, since such a large share of the included studies did not take this attributive step, it was also not viable to exclude non-attributive studies from most of the scenarios. As a result, plausible and conservative assumptions were made as to the relative contribution by CGIAR institutions. If available, the IARC proportion of the total research budget for the innovation was utilized as an attributive percentage. When attribution was not conducted in an included study, and there was no empirical basis on which attribution could easily be assessed, a blanket attribution level of 50% was utilized. This value was selected because this approximate percentage has been reflected in empirical assessments of the catalytic value of IARC efforts to the total number of outputs produced through NARS-IARC collaboration, as well as in observed average proportions of genetic content in collaborative breeding for major commodities (Evenson and Gollin, 1997; Byerlee and Traxler, 1995). The blanket assumed 50% attribution level offers an implicit counterfactual scenario that in the absence of IARC participation, 50% of the observed research benefit would have been realized.

2.6.2 Deflation and discounting

Once adjustments were made to estimate IARC-attributable benefit levels in U.S. Dollars, nominal deflation/inflation of currency values was calculated via the U.S. Producer Price Index, so as to establish a common base-currency year of 1990 for all included benefits. This was performed independently for each of the studies, as most annual benefit levels were already calculated according to the base currency years of each of the studies, while a few presented annual benefits in nominal values. Once nominally adjusted, benefits from the included studies were aggregated to produce total annual benefit streams, and these total annual benefit streams were discounted using a 2% real social discount rate, with sensitivity analyses lowering the rate to 0% and raising it to 10%. This range of rates was chosen because it represents a realistic range of potential returns to very long-term private-sector alternative investments. While a higher set of rates may be more appropriate for relatively short-term investments, such is a plausible rate of return over several decades or more.

Although the included benefits estimates only cover a small sample of CGIAR activities, they were set against comprehensive cost estimates for every activity of the System, with the benefits from all other actions of the CGIAR omitted. Costs were estimated from total CGIAR System investments reported in *Integrative Reports, Financial Reports* and *Annual Reports* published between 1974 and 2001. For 1995, 1996 and 1997, the money awarded to IFPRI by the Asian Development Bank for its rice policy assistance to Vietnam was also added as a cost, since this activity was externally funded. In addition, rice breeding benefits for Latin America and the global value of modern spring bread wheat varieties included benefits derived from research by IRRI, CIMMYT and CIAT that predated the establishment of the CGIAR. To incorporate the costs of additional early research efforts leading to these benefits, investments in IRRI, CIMMYT and CIAT from 1960, 1966 and 1966, respectively, were added to investments after System establishment. These costs were deflated to a 1990 base year, and then discounted according to a 2% real social discount rate.

To express the aggregation process algebraically:

$$TV_u = \sum_{t=s}^{n} \sum_{i=1}^{z} \frac{B_{it} a_{it}}{(1+r)^t} \qquad TC_u = \sum_{t=f}^{j} \sum_{c=1}^{q} \frac{K_{ct}}{(1+r)^t} \qquad BCR_u = \frac{\sum_{t=s}^{n} \sum_{i=1}^{z} \frac{B_{it} a_{it}}{(1+r)^t}}{\sum_{t=f}^{j} \sum_{c}^{q} \frac{K_{ct}}{(1+r)^t}}$$

TV = total value of benefits assessed
u = scenario under which estimate is generated
t = year (1990, the base year of the study, equals 0)
s = start year of benefit period
n = end year of benefit period
i = particular study included
z = total number of studies reporting benefits for year
B = benefit value reported in study (in 1990 US dollars)
a = attributive coefficient (if B is empirically attributed this equals 1, otherwise 0.5 is used)
r = real discount rate
TC = total costs of CGIAR and related investments
f = first year of IARC investment associated with outputs of the CGIAR
j = most recent year of CGIAR investment
c = IARC receiving investment
q = number of IARCs receiving CGIAR-related investment
K = investments in IARC
BCR = benefit-cost ratio

2.7 Limitations of the B-C Meta-Analysis

The validity and accuracy of the present meta-analysis' approach is contingent upon a number of key assumptions. Perhaps most significant of these is the supposition that the presence of the CGIAR System has not resulted in any significant "poisoned wells," or outputs with significantly negative impacts. Such may or may not be indeed completely accurate, as it is very possible that certain problems, such as the transmission of specific exotic pests, may have been potentially caused through individual System actions. Furthermore, according to Alston *et al.* (2000) "it is more likely that past R&D (particularly private R&D) has led to technologies that exacerbate, rather than ameliorate, the negative environmental consequences of agriculture, so the omission of these effects has given rise to generally overstated social rates of return." However, no systematic effort to-date has attempted to analyse the impacts of unintended or inappropriate outputs within the CGIAR (such as accidental pest introductions), and it is likely that if such mistakes were indeed made, they would be very difficult to accurately attribute to specific actions or actors. There have been attempts to comprehensively analyse negative externalities associated with practices accompanying the adoption of IARC-fostered innovations, but these have only been on a qualitative basis (Maredia and Pingali, 2001). Furthermore, they generally conclude that research, or the generation of new knowledge, is difficult to attribute as the source of negative externalities, since these are largely the products of accompanying practices, not the CGIAR output itself. Due to these problems of attribution and quantitative data availability, it is not within the limited scope of the present analysis to attempt to account for negative impacts.

While the present analysis can offer meaningful insights into minimum aggregate levels of benefits generated through the efforts of the CGIAR, such can by no means be considered comprehensive. Most of the System's impacts have not been subjected to thorough assessment, and many do not lend themselves to easy quantification. Impact assessment at the System level is a relatively young activity. As a consequence, the magnitude, comprehensiveness and methodologies for impact assessment differ significantly among Centres and research activities. Furthermore, there is presently a paucity of methods for quantifying longer-term mandate-level impacts on the mutidimensional problems of poverty, and this is necessary for evaluating true progress towards System goals. Many types of research pursued in the CGIAR also have impact pathways that make attribution especially difficult (such as policy research), or lead to benefits that are difficult to quantify (such as certain kinds of natural resource management research). Therefore, it should be stressed that the absence of quantified benefits for most research areas should not be extrapolated to imply that the impacts generated by such have been insignificant. For this reason, current impact-assessment coverage is insufficient to allow for truly comprehensive analysis or significant allocative insights. Thus, while the present meta-analysis can demonstrate whether past investment in the entire System has been minimally justified by known and measurable benefits, the results should not be used to substantiate or inform future allocation among CGIAR activities. Rather, the significance of the benefit levels generated from such a small sampling of research outputs should offer meaningful proof of the productivity of the overall research portfolio.

Furthermore, while the present study was being conducted, there was only limited opportunity for interaction with the authors of included analyses. Thus, when the methodology of the analysis is not clear in the text of reviewed the studies, assumptions had to be made from the best available evidence presented. In some cases these assumptions may over or underestimate the methodological sophistication of the included analyses.

3. RESULTS

The literature survey, in combination with the criteria for selecting "plausible" analyses, results in a selection of 15 studies for the more comprehensive aggregate benefit values, of which a smaller subset is rated as "significantly demonstrated" (Table 3). Application of the review principles, criteria and indicators to the studies reporting the five largest benefit values is illustrated in Table 6, while key characteristics of all 15 studies are described in Appendix II.

As illustrated by Table 4, only a small sampling of the probable total benefits generated by the CGIAR is represented by even the most comprehensive assortment of studies. The small sample of research programmes included and limited time period covered by the studies results in limited potential for comparative evaluation, so such is not a major emphasis within the present analysis. Cumulative profiles of benefits produced over time under the five scenarios are shown in Figure 8, against the investments in the System, which total $7.12 billion by the end of 2001 (in 1990 dollars). Aggregate annual benefit streams under different scenarios and discount rates are presented in Table 8, while overall benefit-cost ratios are depicted in Figure 9.

Table 3. Numbers of studies included in the five scenarios of the present meta-analysis

	1. Significantly demonstrated & empirically attributed	2. Significantly demonstrated	3. Plausible (no extrapolation)	4. Plausible, extrapolated to the present	4. Plausible, extrapolated through 2011
Number of studies	4	7	15	15	15

1. "Significantly demonstrated & empirically attributed" scenario

To be extremely conservative, and go beyond commonly accepted impact analysis standards, only "significantly demonstrated" analyses that include empirically-derived attribution of CGIAR contributions to impacts may be included in an aggregate estimate of benefits. Application of this very restrictive standard allows four[5] studies to be included, which represent four of the Centres. One of these studies utilized computer modelling techniques, while the remaining three used economic-surplus measures. Collectively, these studies produce a benefit-cost ratio of 1.94 (with a 2% real discount rate), which is substantially higher than unity, and a respectable internal rate of return of 17.29% (Table 7; Figure 10). With no discounting applied this ratio becomes 1.95, and if a 10% rate is used, the ratio produced is 1.56.

This type of estimate is of higher confidence than estimates utilising assumed attributive coefficients, but since attribution by institution has not been a recommended IA practice, not many studies fulfilled this criterion, and several robust analyses were omitted. Still, the likely contributions of collaborating organizations may be underestimated through

[5] Even though not empirically attributed, the study assessing cassava-mealybug biocontrol was included in this scenario, since approximately 80% of the total expenditures for this research programme are borne by IITA, a similar figure of 80% of the total derived benefits are attributed to the Centre.

assumed attributive coefficients, and an empirical basis provides much more solid grounding for assessing CGIAR additionality.

2. "Significantly demonstrated" scenario

The very conservative small sampling of "significantly demonstrated" assessments within the System includes seven studies representing six of the Centres. Two of these studies utilized computer modelling techniques, while the remaining five used economic-surplus measures.

Those studies classified as "plausible" but not "significantly demonstrated" may indeed encompass comparable rigour, but such was not clearly evident in the publication, as methodological details often remained unclear. To be extremely conservative, such opacities were denied "the benefit of the doubt," and studies without clearly presented analytical details, such as survey sample sizes, or types of data sources for productivity estimates were excluded from the "significantly demonstrated" scenario. Where methodological details were more apparent, a few studies were excluded on the basis of potential significant subjective bias in key data sources, such as estimates of adoption or productivity impacts based chiefly on the testimony of a single expert or on a very small number of beneficiary interviews. The methodological sophistication of "significantly demonstrated" studies is higher, in general, than is the sophistication of "plausible" studies, as productivity impacts were disaggregated by agro-ecological conditions in all of these studies, while few "plausible" studies took this measure.

The bulk of values reported in this scenario have been generated by three key innovations – modern varieties of rice, modern varieties of wheat and cassava-mealybug biocontrol, as such comprise 98.1% of "significantly demonstrated" benefits. After discounting and nominal adjustment, this scenario produced a benefit cost ratio of 3.77, which is considerably higher than unity. Sensitivity analyses that raise the social discount rate to 10% and lower it to zero result in ratios of 3.40 and 3.72, respectively[6]. Even with this highly conservative sampling and sensitivity analysis, the ratio produced is a substantial multiple of unity, and the IRR of 33.91% is considerable.

3. "Plausible" scenario

Within the broader "plausible" assortment of 15 studies, 11 focussed on crop genetic improvement, two on biological control of pests, one on policy guidance and one on seed production technology. Nine of the 16 IARCs are represented in the selected studies. Twelve studies utilized simple economic-surplus measures, while two used simulation software, and a final study used econometric techniques. All but two of the studies have been or are in the process of being published in peer-reviewed scientific books or journals. The assortment of studies included under this standard appears in Table 4, and additional details on the studies producing the five largest benefit values are presented in Table 5.

[6] Due to the specific distribution of benefits and costs within this scenario, a small discount rate actually maximizes the ratio of benefits to costs.

Table 4. "Plausible" aggregate benefit estimates and periods of coverage within the scope of the CGIAR research agenda. Black shading indicates relatively comprehensive geographic coverage, while grey indicates partial coverage for values included. Note that the reported values have been adjusted for attribution, deflated and discounted (2% real rate) to a 1990 base year.

Research Emphasis Coverage by Large-Scale Plausible Impact Analyses

		1970	1975	1980	1985	1990	1995	2000	Total Benefit (billion $)
Commodities	Barley		Hassan et al., 2002						0.33
	Beans		Johnson et al., 2002						0.59
	Cassava			Johnson et al., 2002					0.23
	Chickpea								
	Coconut								
	Cowpea								
	Groundnut								
	Lentil								
	Maize						1		0.44
	Millet								
	Pigeonpea				Bantilan & Joshi, 1996				0.22
	Plantain								
	Potato								
	Rice - IRRI	Hossain, 1998					2		3.21/4.31
	Rice - CIAT	Sanint & Wood, 1998							8.28
	Rice - WARDA						3		0.15
	Sorghum								
	Soybean								
	Sweetpotato					4			0.25
	Wheat - durum						5		
	Wheat - spring bread		Byerlee & Traxler, 1995				5		9.75/0.88
	Yam								
Livestock	Disease control								
	Genetic improvement								
	Pasture/forages								
Fisheries	Aquaculture								
	Coastal management								
	Genetic improvement								
Natural Resource Management	Crop management								
	Nutrient management								
	Watershed management								
	Biocontrol - cassava mealybug			Zeddies et al., 2001					5.10
	Biocontrol - mango mealybug					6			0.11
Forestry	Forest management								
	Non-timber forest products								
	Plantations								
Biodiversity	Agricultural								
	Aquatic								
	Forest								
Capacity Building	Training								
Policy	Agricultural								
	Economic								
	Economic - rice policy						7		0.04
	Food								
	Forests								
	Livestock								

1. Morris et al., 2001; 2. Hossain et al., 2003; 3. Dalton & Guei, 1998; 4. Fuglie et al., 1997; 5. Heisey et al., 1999; 6. Bokonon-Ganta et al., 2002; 7. Ryan, 1999

From the included assessments, it is clear that spring bread wheat breeding, rice genetic improvement and cassava-mealybug biocontrol have had particularly high research payoffs, as these three research programmes have collectively generated 93.4% of this wider array of estimated benefits (Figure 7). Although comparisons are not possible due to the lack of analyses for other programmes, it is clear that research in these three areas had pervasive and large-scale impacts on wide regions of the globe.

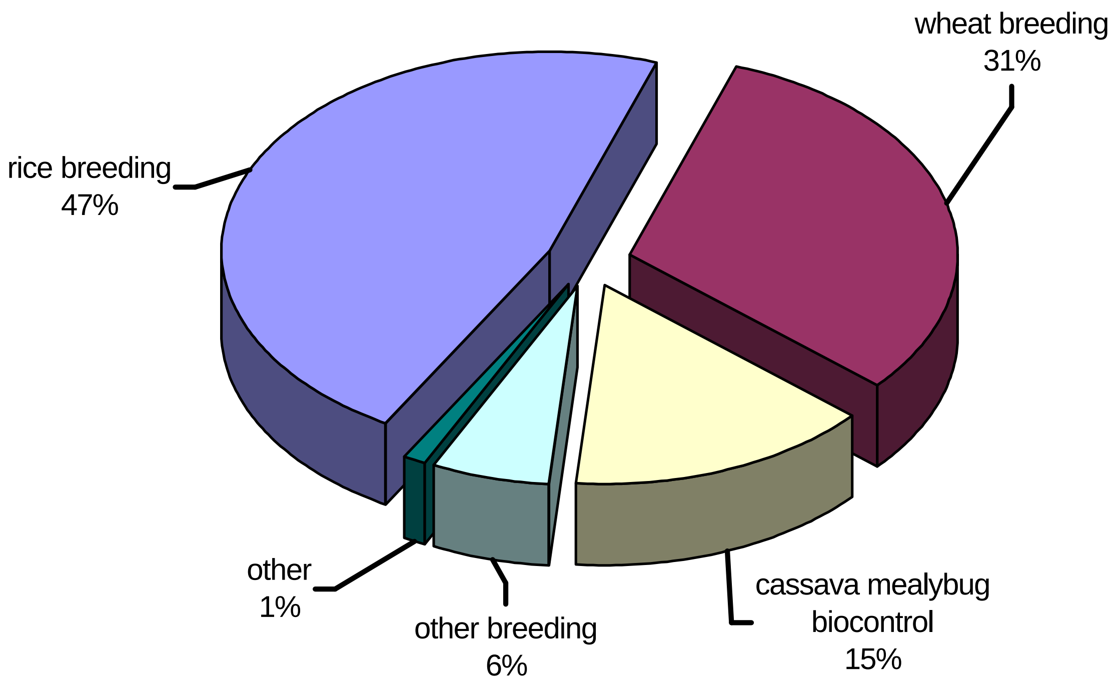


Figure 7. Percentage of benefits derived from different research areas in the scenario of "plausible" studies, with no extrapolation of benefit values reported

In scenario 3, the benefit-cost ratio produced with a 2% real discount rate and nominal adjustment is 4.76 (Figure 9), with a rate of return of 34.01%. This value is considerable according to standard investment analysis. Raising the real discount rate to 10% results in a B-C ratio of 3.78, while lowering it to zero results in 4.91, indicating only modest sensitivity to the discount rate applied.

Table 5. Characteristics of the studies producing the five highest benefit values in the B-C Meta-Analysis. Collectively, these studies account for 90.4% of benefits under the "plausible" scenario (with no extrapolation).

Category of Innovation	Citation	Region	Period	Type of Analysis	Adoption Data Sources	Productivity Data Sources	Attribution	Prices Utilized	Adjusted "Plausible" Benefits (billions of 1990 $)
Cassava-mealbug biocontrol	Zeddies, J., R.P. Schaab, P. Neuenschwander and H.R. Herren. 2001. Economics of biological control of cassava mealybug in Africa. Agricultural Economics. 24: 209-219.	Africa	1974-2013	Economic surplus with disaggregation into 3 "ecological zones"	"Adoption" not a factor, as control agent (parasitic wasp) dispersed naturally, estimates of proliferation and spread based on entomological studies	Crop loss reductions to stable equilibria (return to pre-invasion productivity levels) for the 3 "ecological zones," based on 6 entomological studies; baseline cassava production data sourced from FAO, African governments and IITA's Collaborative Study on Cassava in Africa	No attribution in study, but IITA efforts comprised 80% of total expenditures	World market prices for cassava are utilized for scenario used in the present study	5.10
Modern varieties of rice	Hossain, M., D. Gollin, V. Cabanilla, E. Cabrera, N. Johnson, G.S. Khush and G. McLaren. 2003. International Research and Genetic Improvement in Rice: Evidence from Asia and Latin America. *In* R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Asia	1998	Economic surplus with national disaggregation	National adoption estimates are sourced from the World Rice Statistics Database (no metadata concerning the database are provided in the study)	Costs and return data have been collected by NARS and IRRI during the late 1990s through household surveys (no metadata or citations are provided for the surveys)	No attribution in study	Domestic prices	4.31
Modern varieties of rice	Hossian, M. 1998. Rice research, technical progress, and the impact on the rural economy: the Bangladesh case. *In* Pingali, P.L. and M. Hossain (eds.) Impact of Rice Research. Manila, Philippines: International Rice Research Instiute.	Bangladesh	1973-1993	Economic surplus	Review of 3 published studies and unpublished survey data from national research institutes of Bangladesh (no metadata provided on data sources for cited studies)	Same sources as for adoption - review of three published studies and unpublished survey data from national research institutes of Bangladesh	No attribution in study	Domestic prices for unit cost reductions	3.21
Modern varieties of rice	Sanint, L.R. and S. Wood. 1998. Impact of Rice Research in Latin America and the Caribbean During the Past Three Decades. *In* Pingali, P.L. and M. Hossain (eds.) Impact of Rice Research. Manila, Philippines: International Rice Research Institute.	Latin America and Caribbean	1967-1995	Application of DREAM multimarket economic-surplus model	Review of 10 previously published regional rice studies for Latin America (no metadata provided on data sources for cited studies)	Basic productivity data are derived from the 10 regional studies used for adoption estimates, of which exogenous sources of yield growth are independently specified, based on the same data set.	No attribution in study	Domestic prices for unit cost reductions	8.28
Modern varieties of Spring bread wheat	Byerlee, D. and G. Traxler. 1995. National and International Wheat Improvement Research in the Post-Green Revolution Period: Evolution and Impact. American Journal of Agricultural Economics 77: 268-278.	Global	1966-1990	Economic surplus with disaggregation into 4 "mega-environments" and 4 regions	Utilized the CIMMYT Global Wheat Impact Study, which comprised "annual government surveys in some countries, special surveys at a regional or country level, seed sales in some countries and wheat researchers' estimates."	Yield gain was extrapolated from review of relative gains reported in 12 published studies and 7 unpublished sources (no metadata other than location are provided on each of the cited estimates), and was broken down into Stage 1 (traditional to modern variety) and Stage 2 (modern variety to newer modern variety) gains for "mega-environments" by region.	Based on percentage of CIMMYT derived germplasm (0.85 for CIMMYT crosses, 0.50 for NARS' crosses with CIMMYT parent, etc.)	CIF for net importing regions, FOB for net exporting regions, with world price numeraire	9.75

Table 6. Basis of ratings for the five studies producing the highest benefit values in the B-C Meta-Analysis. Collectively, these studies account for 90.4% of benefits under the "plausible" scenario (with no extrapolation)

Principle	Criteria	Indicator	Low rating	High rating	Byerlee, D., and G. Traxler. 1995. National and International Wheat Improvement Research in the Post-Green Revolution Period: Evolution and Impact. American Journal of Agricultural Economics 77: 268-278.	Hossian, M. 1998. Rice research, technical progress, and the impact on the rural economy: the Bangladesh case. *In* Pingali, P.L. and M. Hossain (eds.) Impact of Rice Research. Manila, Philippines: International Rice Research Institute.	Hossain, M., D. Gollin, V. Cabanilla, E. Cabrera, N. Johnson, G.S. Khush, and G. McLaren. 2003. International Research and Genetic Improvement in Rice: Evidence from Asia and Latin America. *In* R.E. Evenson. Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI	Sanint, L.R. and S. Wood. 1998. Impact of Rice Research in Latin America and the Caribbean During the Past Three Decades. *In* Pingali, P.L. and M. Hossain (eds.) Impact of Rice Research. Manila, Philippines: International Rice Research Institute.	Zeddies, J., R.P. Schaab, P. Neuenschwander, and H.R. Herren. 2001. Economics of biological control of cassava mealybug in Africa. Agricultural Economics. 24: 209-219.
Transparency	apparent basis of key assumptions	explicitness of key assumptions	major assumptions underlying analysis are not defined	all major assumptions explicitly stated	assumptions relatively explicit	most assumptions are relatively explicit	most assumptions are relatively explicit	most assumptions are relatively explicit	most assumptions are relatively explicit
		substantiation of key assumptions	explicit assumptions have no clear basis	explicit assumptions have logical justification and/or citation	relatively well substantiated assumptions	most assumptions are substantiated	most assumptions are substantiated	most assumptions are substantiated	most assumptions are substantiated
	complete citation of data sources	citation of adoption data	unclear basis of adoption estimates	adoption estimates cited and/or methodology described	methodology is clearly explained, with a sample survey questionnaire	adoption data are cited from published and unpublished sources, but not described in derivation	adoption data are cited but not described in derivation	reasonably well cited	adoption not relevant as control agent dispersed naturally
		citation of productivity data	unclear basis for productivity claims	productivity claims based on cited references or clear methods	studies used for calculating yield advantage are clearly cited, although no metadata are provided	productivity data are cited from published and unpublished sources, but not described in derivation	productivity data are cited from unpublished studies and are not described in derivation	reasonably well cited	reasonably well cited
		citation of adoption-related costs data	unclear empirical basis for deriving costs associated with adoption	estimates of adoption-related costs (where considered) cited or given logical justification	somewhat unclear	adoption-related costs data are cited from published and unpublished sources, but not described in derivation	adoption-related cost data are cited from unpublished studies and are not described in derivation	somewhat unclear	no significant adoption-related costs
		citation of price sources	unexplained basis of commodity prices	cited basis for commodity prices	prices are clearly cited	unclear	unclear	somewhat unclear	prices are clearly cited
	full explanation of data treatment	explanation of scaling-up adoption estimates	no basis provided for adoption estimates	gathering process for adoption estimates defined	fairy clearly explained.	unclear	fairly clear	somewhat unclear	NA
		explanation of scaling-up productivity estimates	unclear extrapolation from limited productivity impact data	clearly defined methodology for scaling-up site-specific impact estimates	somewhat clear, although it appears that fairly few studies are used for estimating global "Stage One" yield gains	fairly clear	somewhat unclear	somewhat unclear	fairly clear
		explanation of scaling-up adoption-related costs	unclear manner of incorporation of costs associated with adoption	costs considered (or not considered) in an explicit manner	not clearly explained in the study.	fairly clear	fairly clear	somewhat unclear	NA
		explanation of economic valuation	commodity prices used, discounting and deflating unclear	commodity prices used, discounting and deflating clearly presented	very clear	somewhat clear	clear	clear	clear
Demonstration of causality	representative data set utilised	reliability of data utilised	data sourced from uncorroborated expert opinion	data sourced from empirical studies or methods and validated through triangulation	fairly reliable data - probably as good as can be expected	the data used appear fairly robust	difficult to assess, as few details are provided about data sources	the data used appear to be robust	FAO data for cassava production, which appear to be extensively used, are notoriously unreliable, but there's no better alternative
		comprehensiveness of data set utilised	data sourced from a small set of unrepresentative sites	large number of sample sites representing the range of relevant agro-environments.	very comprehensive for Stage II, less clearly so for Stage I	it appears that the data set utilised is broadly representative at a national level	difficult to assess, as few details are provided about data sources	appears comprehensive, as this study had many earlier impact assessments up which it could rely	many studies are cited to substantiate the spread and effects of the control agent
	adequate consideration of mitigating influences	appropriateness and completeness of analysis of mitigating factors	no mitigating factors considered	comprehensive consideration of all major relevant alternative causal factors	the study could benefit from greater consideration of substitution possibilities, but is otherwise adequate	fairly complete analysis of mitigating factors and adoption context	mitigating factors are not extensively considered	model used appears to encompass a number of potential mitigating factors	appears adequate, due to the clear attributability of the biocontrol programme
	appropriate disaggregation	disaggregation by production environment	only "average" conditions considered	heterogeneity in impacts appropriately captured	disaggregation by mega-environment and region - appropriate	discussion of differential adoption by different production environments	national disaggregation - it would be more illustrative to disaggregation by production environment (paddy, upland, etc)	disaggregation into four production environments - appropriate	three "ecological zones" are utilised, and this appears adequate
		consumer vs. producer surplus	gross benefits presented without analysis of surplus recipients	impacts disaggregated among different producer and consumer groups	no disaggregation of surplus generated between producers and consumers	no disaggregation of surplus generated between producers and consumers	no disaggregation of surplus generated between producers and consumers	disaggregation between four production methods and consumers	no disaggregation of surplus generated between producers and consumers
	plausible counterfactual developed	explicitness of counterfactual	no "without scenario" presented	without scenario comprehensively developed	counterfactual is somewhat inexplicit	explicit counterfactual	implicit counterfactual	explicit counterfactual	four explicit counterfactual scenarios are developed
		counterfactual plausibility	counterfactual represents unrealistic, overly cynical, course of action	counterfactual represents realistic, likely, and substantiated path of events	implicit counterfactual is highly plausible	plausible counterfactual based on the yield potential of traditional varieties	reasonably plausible counterfactual	reasonably plausible counterfactual	the counterfactuals appear somewhat over-pessimistic as it seems likely that farmers would find some better strategy than living with 40% of traditional cassava yields
	precise institutional attribution	plausibility of attribution	no attribution attempted	empirically-based attribution derived through counterfactual	attribution is based on CIMMYT-derived germplasm content	no attribution attempted	no attribution attempted	no attribution attempted	no attribution attempted
Classification					Significantly demonstrated	Significantly demonstrated	Plausible	Significantly demonstrated	Significantly demonstrated

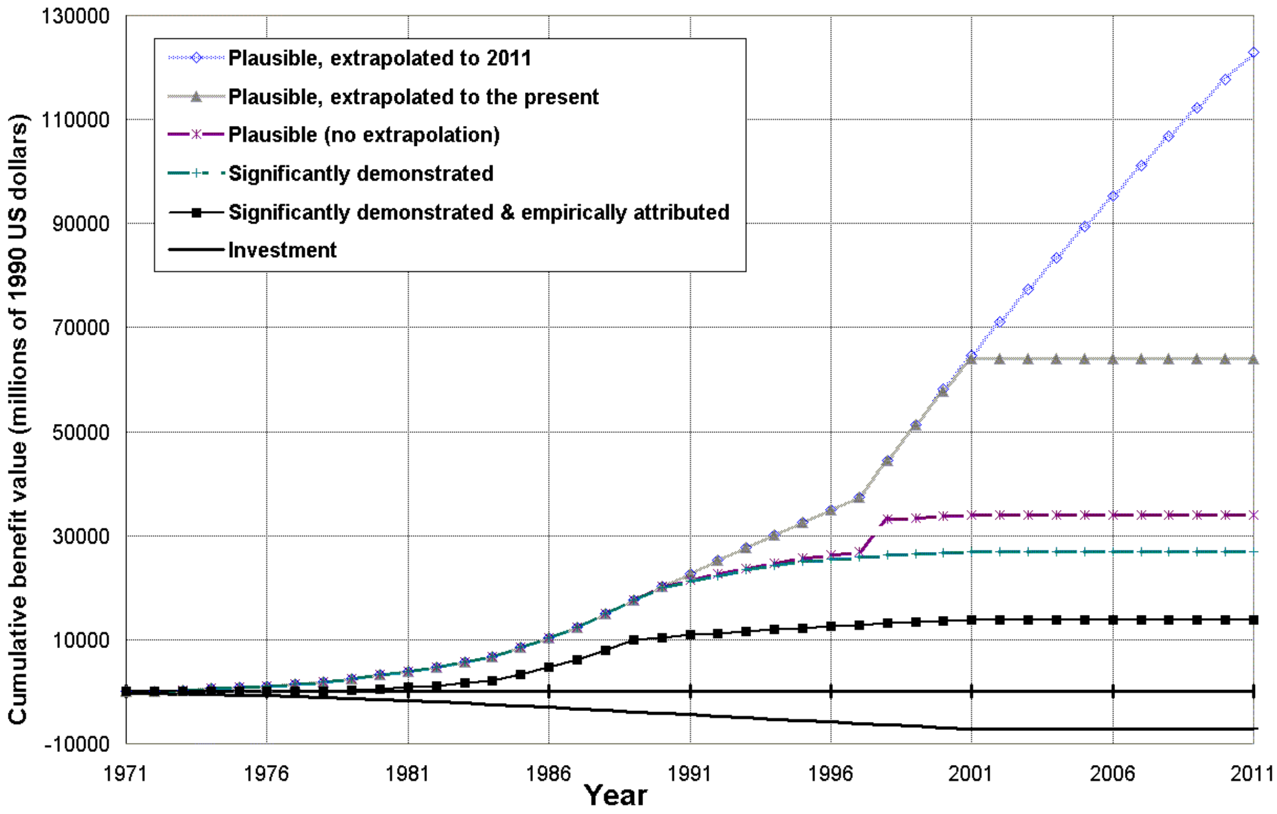


Figure 8. Cumulative estimates of benefits from and investments by the Consultative Group on International Agricultural Research in the activities of the International Agricultural Research Centres over time, according to different scenarios of study selection

4. "Plausible, extrapolated to the present" scenario

The benefits derived could also be plausibly extrapolated under the assumption that single-year benefit estimates for modern varieties also represent years immediately subsequent to the year of the estimate (i.e. a benefit estimate based on 1998 data may also apply to 1999, 2000 and 2001). Furthermore, benefit estimates over many years representing a progressively rising trend until the terminal year of the assessment, may be conservatively assumed to continue at the benefit level of the final year until the present. Benefits derived, particularly for modern varieties, cannot be expected to disappear after the end of the estimated period, as germplasm has a useful lifespan of at least a decade, even if no further improvements are made (germplasm improvement constitutes the bulk of estimated benefits). Furthermore, rising benefit trends for all of the multiyear genetic improvement studies indicate that productivity gains continue to be made, so even this may be regarded as a conservative conjecture. Under these assumptions, the benefit-cost ratio rises to 9.00, with an IRR of 34.41%. With no discounting applied, the ratio is 10.03, and if a 10% discount rate is utilized, the ratio becomes 5.40.

5. "Plausible, extrapolated through 2011" scenario

To go further, presently conducted research has a significant lag period before benefits are generated, as the core IARC innovation must be locally adapted by NARS, and often must be approved by national governments before widespread dissemination is pursued.

Furthermore, the innovation must then have time to diffuse among farmers and be adopted before benefits are generated. In general, this process usually takes at least one decade. Thus, the bulk of benefits measured in these studies were generated by research that happened at least a decade ago, and most research conducted in the present will not have impact until at least ten years from now. Therefore, it can be assumed that if returns to research conducted in these specific areas have been at least constant over the past decade, benefits could continue unabated until at least ten years into the future, if research stopped at present.

Such continued benefits may be considered as a reasonable, although somewhat speculative assumption, based on observations regarding the continued relevance of assessed innovations. The impact of an innovation persists until the innovation becomes obsolete or replaced by antecedent alternative technology (Alston *et al.*, 1998a). Accordingly, it can be plausibly expected that semi-dwarfism and disease resistance will not become obsolete in the next decade. For instance, adoption of IARC derived varieties of wheat and rice, which comprise most of the assessed benefits, has continued to rise in recent years, and shows no sign of decreasing, according to the reviewed literature (Heisey *et al.*, 2003; Hossain *et al.*, 2003). Furthermore, yield gains from past research have largely been maintained or expanded through second and third generation varietal replacement, which have fostered increased disease resistance (Byerlee and Traxler, 1995). In addition, biocontrol of the cassava-mealybug has been predicted to remain viable until at least 2013 (Zeddies *et al.*, 2001). However, it should be stressed that while this assumption may be regarded as a more complete estimation of benefits from investments to-date, such is a very tenuous and rough approximation of potential events.

If this speculative assumption is combined with application of the expected future results of cassava mealybug biocontrol, the benefit-cost ratio rises to an outstanding 17.26, with 122.9 billion (1990) dollars of benefits generated (Figure 8). When the real discount rate is raised to 10%, this ratio falls to 7.07, and with no discounting it rises to 21.83. The internal rate of return produced by these benefits is not much higher than the other scenarios at 34.48%, due to the fact that most benefits occur long after investments.

Table 7. Internal rates of return produced by benefit values aggregated from large-scale economic impact assessments of CGIAR research, under five scenarios of study selection

1. Significantly demonstrated & empirically attributed	2. Significantly demonstrated	3. Plausible (no extrapolation)	4. Plausible, extrapolated to the present	4. Plausible, extrapolated through 2011
17.29%	33.91%	34.01%	34.41%	34.48%

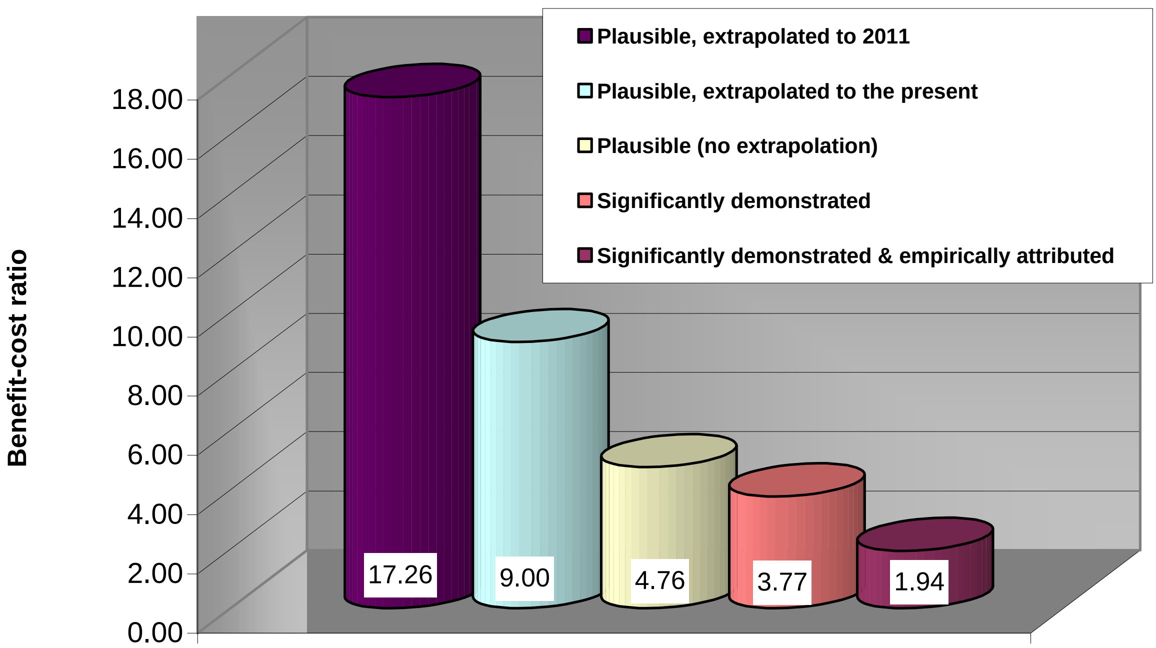


Figure 9. Aggregate ratios of benefits derived from research and investment by the Consultative Group on International Agricultural Research in the activities of the International Agricultural Research Centres, as compared with costs, over time, under different scenarios of study selection

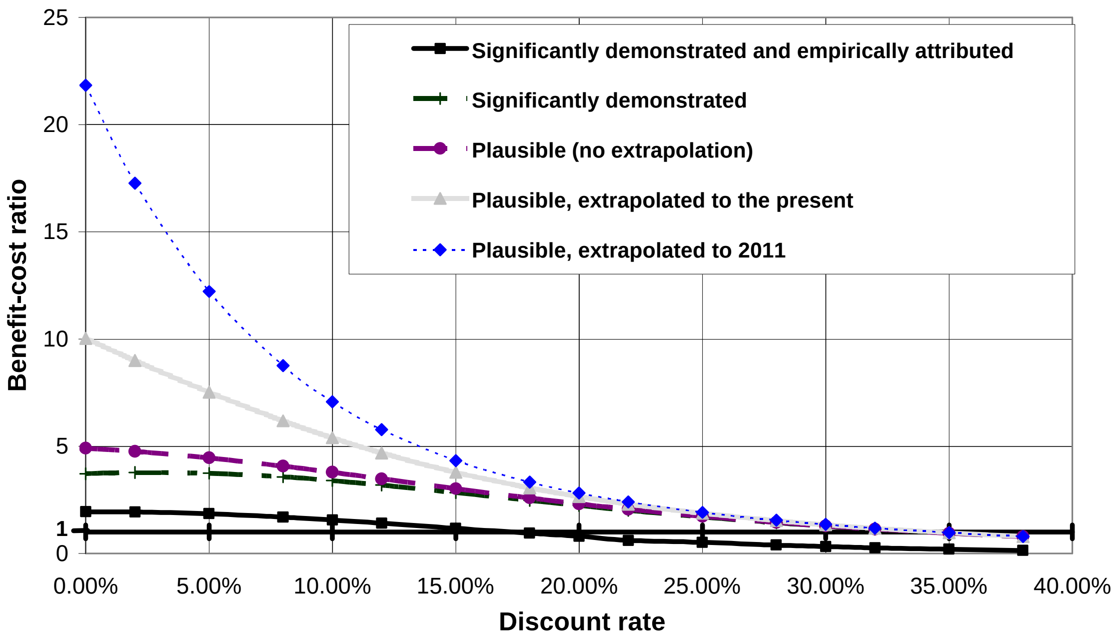


Figure 10. Effect of discount rate on aggregate benefit-cost ratios produced under different scenarios of study selection

Table 8. Table of benefit estimates aggregated from values reported in large-scale economic impact studies of CGIAR research, under different scenarios of study selection and interest rates (all values in millions of 1990 US dollars)

Year	Costs			Significantly demonstrated & empirically attributed			Significantly demonstrated			Plausible (no extrapolation)			Plausible, extrapolated to the present		
	2% real rate	no discounting	10% real rate	2% real rate	no discounting	10% real rate	2% real rate	no discounting	10% real rate	2% real rate	no discounting	10% real rate	2% real rate	no discounting	10% real rate
1960	1.06	0.59	10.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1961	1.50	0.84	13.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1962	7.39	4.24	61.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1963	3.44	2.01	26.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1964	4.22	2.52	30.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1965	14.57	8.88	96.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1966	12.54	7.79	76.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1967	38.48	24.40	218.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1968	25.59	16.55	134.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1969	40.80	26.92	199.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1970	40.15	27.02	181.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1971	65.99	45.30	277.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1972	83.73	58.63	325.95	0.00	0.00	0.00	112.16	78.53	436.62	112.16	78.53	436.62	112.16	78.53	436.62
1973	90.25	64.45	325.78	0.00	0.00	0.00	226.46	161.73	817.44	226.46	161.73	817.44	226.46	161.73	817.44
1974	103.05	75.07	344.94	0.00	0.00	0.00	255.95	186.45	856.72	255.95	186.45	856.72	255.95	186.45	856.72
1975	127.47	94.71	395.65	0.00	0.00	0.00	291.14	216.32	903.64	291.14	216.32	903.64	291.14	216.32	903.64
1976	158.06	119.79	454.90	0.00	0.00	0.00	287.87	218.17	828.51	287.87	218.17	828.51	287.87	218.17	828.51
1977	183.76	142.05	490.41	0.00	0.00	0.00	316.77	244.87	845.37	316.77	244.87	845.37	316.77	244.87	845.37
1978	179.83	141.79	445.01	72.88	57.47	180.35	434.03	342.23	1074.07	434.03	342.23	1074.07	434.03	342.23	1074.07
1979	180.98	145.55	415.28	205.97	165.66	472.64	593.31	477.18	1361.44	593.34	477.20	1361.51	593.34	477.20	1361.51
1980	188.75	154.84	401.61	276.18	226.57	587.66	711.96	584.06	1514.90	712.03	584.11	1515.04	712.03	584.11	1515.04
1981	185.58	155.28	366.15	284.54	238.09	561.41	738.36	617.82	1456.80	738.49	617.93	1457.05	738.49	617.93	1457.05
1982	234.05	199.76	428.19	321.11	274.06	587.48	800.99	683.63	1465.43	801.46	684.04	1466.30	801.46	684.04	1466.30
1983	248.43	216.27	421.46	497.86	433.42	844.61	1005.86	875.66	1706.42	1007.19	876.82	1708.67	1007.19	876.82	1708.67
1984	256.43	227.70	403.39	532.59	472.92	837.81	1043.71	926.79	1641.86	1046.57	929.32	1646.35	1046.57	929.32	1646.35
1985	261.15	236.53	380.93	1190.24	1078.03	1736.18	1724.18	1561.65	2515.05	1732.74	1569.40	2527.53	1732.74	1569.40	2527.53
1986	264.53	244.39	357.81	1294.96	1196.35	1751.57	1839.20	1699.14	2487.71	1854.08	1712.88	2507.83	1854.08	1712.88	2507.83
1987	267.57	252.13	335.59	1453.71	1369.86	1823.28	2039.66	1922.02	2558.20	2054.31	1935.82	2576.58	2054.31	1935.82	2576.58
1988	295.63	284.15	343.82	1908.26	1834.16	2219.34	2516.01	2418.31	2926.15	2538.94	2440.35	2952.82	2538.94	2440.35	2952.82
1989	287.24	281.61	309.77	1966.83	1928.26	2121.09	2644.06	2592.21	2851.43	2670.07	2617.72	2879.49	2670.07	2617.72	2879.49
1990	287.90	287.90	287.90	418.63	418.63	418.63	2450.82	2450.82	2450.82	2494.79	2494.79	2494.79	2494.79	2494.79	2494.79
1991	284.70	290.40	264.00	463.43	472.70	429.72	1179.03	1202.61	1093.28	1242.01	1266.85	1151.68	2560.15	2611.36	2373.96
1992	303.97	316.25	261.37	375.65	390.82	323.00	1130.91	1176.59	972.39	1186.67	1234.61	1020.34	2478.97	2579.12	2131.50
1993	286.89	304.45	228.74	332.62	352.98	265.20	1090.54	1157.29	869.49	1174.76	1246.67	936.64	2441.72	2591.17	1946.79
1994	290.06	313.97	214.45	323.52	350.19	239.18	869.37	941.03	642.74	996.28	1078.41	736.57	2462.02	2664.97	1820.21
1995	277.59	306.49	190.30	285.29	314.99	195.58	847.34	935.53	580.89	999.68	1103.73	685.33	2436.67	2690.28	1670.45
1996	273.36	307.84	173.77	311.92	351.27	198.28	331.89	373.77	210.98	496.50	559.13	315.62	2436.36	2743.74	1548.77
1997	264.32	303.62	155.80	350.47	402.58	206.58	370.05	425.07	218.13	556.75	639.53	328.18	2458.58	2824.14	1449.23
1998	271.20	317.75	148.23	347.71	407.39	190.05	366.91	429.89	200.55	6329.06	7415.51	3459.39	7109.60	8330.03	3886.02
1999	255.92	305.85	129.71	210.05	251.02	106.46	228.87	273.52	116.00	278.40	332.72	141.10	6804.53	8132.05	3448.78
2000	237.89	289.98	111.80	195.16	237.89	91.72	213.61	260.39	100.39	267.32	325.87	125.64	6433.70	7842.64	3023.68
2001	234.19	291.18	102.06	181.39	225.54	79.05	199.48	248.03	86.93	203.21	252.67	88.56	6256.04	7778.60	2726.35
2002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2004	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2006	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2007	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2008	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2009	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2010	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2011	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SUM	7120.20	6897.47	10540.32	13800.96	13450.85	16466.88	26860.52	25681.33	35790.35	33899.05	33844.38	39845.37	64046.74	69176.78	56882.06

4. DISCUSSION

4.1 Effectiveness of the CGIAR investment

4.1.1 Accuracy of benefit estimates

All scenarios produced significant net benefits and benefit-cost ratios well above one, even when a high discount rate is used and when extremely conservative criteria are applied, which restricted the included studies to a select few. This strongly suggests that the CGIAR has been a productive investment, and demonstrates a certain degree of robustness in the results. Sensitivity analyses, using a 10% and zero real discount rates, in addition to nominal adjustments, also maintain this result. The insensitivity to discount rate applied may initially appear counterintuitive, but is largely a product of the benefit distribution peaking during the middle of the period for the non-extrapolative scenarios, with costs evenly spread over the benefit duration (Figure 8).

With limited resources, impact assessors can either choose to invest in enhanced precision and detailed coverage of a specific and limited locality, or they can attempt to encompass larger temporal and spatial scales, but with lower reliability and significantly more reliance upon simplifying assumptions. Since the studies reviewed in the present analysis are all of large scope, all of these studies to some degree sacrifice precision in favour of scale. Encompassing greater scale allows for a more complete picture of impact magnitude, but by nature of lower precision, such estimates are more prone to error, and are vulnerable to inaccurate and unsubstantiated assumptions (Figure 11). This comprehensiveness-precision tradeoff applies both to individual studies included within the present analysis and to scenarios developed within the present study. The more comprehensive extrapolative scenarios encompass more of the “true picture” but do so with considerably less reliability than the most restrictive scenarios. To be sure, the application of different minimum standards for impact demonstration greatly affects results, as the most uncertain scenario produces a benefit value nearly nine times as great as the most conservative. Unless minimum standards for acceptance of impact claims are defined, it is difficult to select one of the scenarios for future reference. However, even the most comprehensive scenario included in the present study probably doesn’t encompass most of the System’s impacts, as temporal and spatial coverages are often not complete, and many acknowledged impacts have never been comprehensively assessed.

This limited IA coverage renders it likely that even the most generous of these results may be considered as somewhat conservative. Accordingly, the “significantly demonstrated” scenarios, though more conservative than the “plausible” scenarios, do not produce more accurate values, but represent a viable absolute minimum level of impact. The extrapolative “plausible” scenarios are probably more accurate (closer to the truth), but are less precise (less repeatable), as the methods applied to derive these values are presented in less detail, and rely more upon assumption.

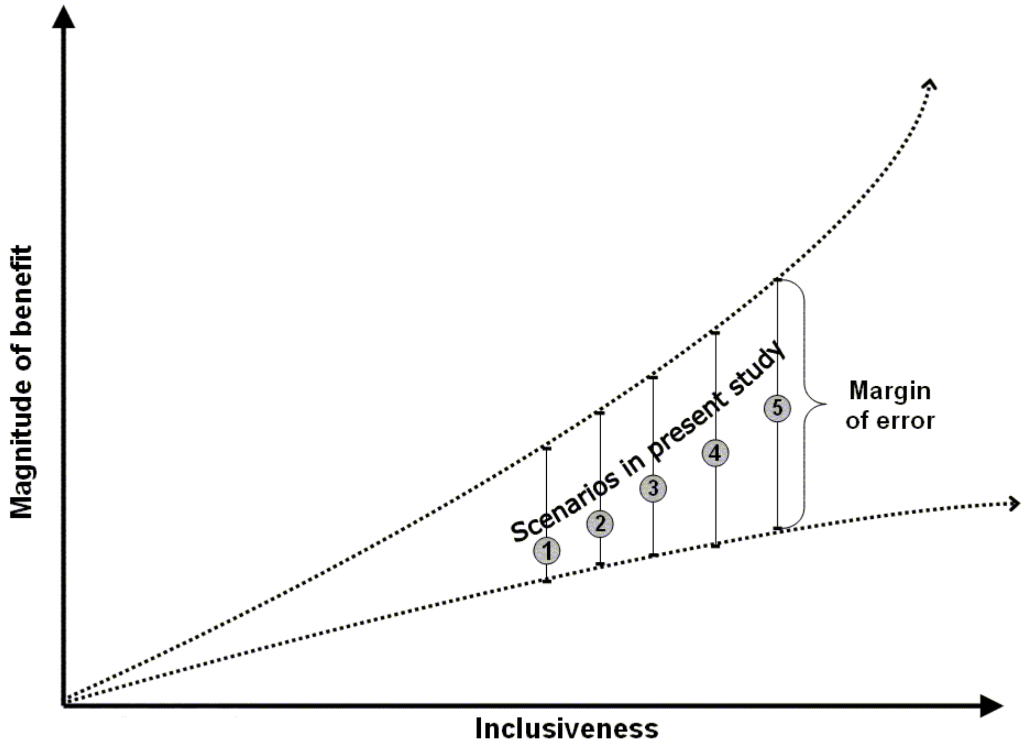


Figure 11. Relationship between inclusiveness of benefit coverage and the error margins of estimates made

Although these values are quite robust, there is potential for over-estimation, even though this is very likely to be outweighed by the many other unquantified CGIAR-derived benefits omitted from the selection of analyses included. Over-estimation may have resulted from the inclusion of benefits derived from innovations that pre-date the establishment of the CGIAR. Modern semi-dwarf varieties of wheat and rice, which comprise the largest documented System research impacts, both had been well established before 1971. Thus, it is arguable that research adoption lags render many of the included benefits attributable to research pre-dating the CGIAR institutions. However, by including investments dating as far back as 1960 in Centres which later became affiliated with the CGIAR System, this potential pitfall has been somewhat addressed, unless research lags are assumed to be very long. The issue of accurately incorporating adoption lags has been long contended in impact assessment, and has great ramifications for analytical results, in terms of both magnitude and attribution. Accordingly, Alston *et al.*, (1998a), have noted that rates of return to agricultural research fall from exceptionally high to comparable with other investments when an infinite lag structure is utilized. Yet, such potential criticism does not substantially apply with regard to the present study, because this study focuses on the marginal effects of applied research, which would not have occurred without the additionality of the assessed activities, even if the scientific basis for such is largely derived from previous innovations. Had the additional efforts of the CGIAR not been pursued, the vast majority of benefits enjoyed would indeed not have been realized, as the pre-CGIAR modern varieties are vastly outyielded by their descendants, which were derived as a result of CGIAR investment. Although adoption lags do present problems in temporal attribution of total effects, marginal additionality of specific actions is largely unaffected.

A second major potential source of error relates to the attribution of simultaneous complementary efforts by non-CGIAR institutions. Most of the included studies, even within the "significantly demonstrated" scenario, do not attempt to partition benefits to different entities, and assessed collaborative efforts holistically. This renders impact claims of specific institutions somewhat tenuous, as there is no empirical basis for assessing the additionality of particular efforts. Furthermore, even when empirical attribution is attempted, the rationale for such is somewhat simplistic, and based on fairly arbitrary assumptions regarding the contribution of genetic content, or reductions in research lag periods. Important contributory factors, including extension, are often not considered, although extension is liable to be of lower significance for the research programmes generating the bulk of assessed benefits (breeding and biocontrol) than for other areas of emphasis (such as crop management). Since the assumed attributive coefficients applied in the present study may have been too generous towards CGIAR institutions, these assumptions may be made even more conservative through further reductions in derived benefit values. To account for such potential for overestimation, if all benefit values reported in the "significantly demonstrated" scenario are halved, the benefit cost ratio still remains reasonably high at 1.88. In fact, if only 27% of the "significantly demonstrated" benefits or slightly more than half of the "empirically attributed" benefits have actually been realized as a result of CGIAR activity, the investment remains sound.

Since most of the reviewed studies use financial prices, without accounting for market distortions, there may be a third significant source of bias if significant social costs accompany these benefits. Indeed, there have been substantial interventions in most agricultural markets of both developing and developed countries during the period assessed. Consequently, there may be substantial scope for over-estimation of returns if significant government costs producing few benefits accompany each unit of production, such as occurs with price-support systems and output subsidies under conditions of surplus production (Oehmke, 1988). Although such interventions have been prevalent in most of the developed world (Western Europe, Japan and North America) during the analysed period, market distortions in much of the developing world are of a distinctly different character. In these countries agriculture was often penalized, rather than subsidized, through various combinations of explicit and implicit taxes. For example, monopsonistic government distribution and marketing agents often bought a significant share of production at below-market prices to offer cheaper food for sale to politically preferred clientele (Sran and Srinivasan, 1987). Export taxes, restrictions or quotas often accompanied these measures, to insure insulation of artificially-depressed prices from world markets, and, perhaps more significantly, macroeconomic policies, such as overvalued exchange rates, essentially imposed high levels of implicit taxation (Binswanger and Deininger, 1997). To compensate for the production-inhibiting consequences of such policies, inputs were often subsidized, and imports were restricted. Overall, agriculture, as a major portion of the economy, was often a net source of government revenue for investment in other sectors, rather than a net sink, as in developed countries (Bates, 1983). Under such circumstances, productivity increases, which raise the quantities of food supplied, and reduce market price equilibria, are less likely to increase government expenditures, although government revenues may decline under certain circumstances. However, from the mid-1980s onward, Structural Adjustment Programmes encouraged by the international donor community, along with ongoing processes of international trade liberalization, have fostered widespread reduction of such market interventions. Under these expanding conditions of liberalization, which should be prevalent for the foreseeable future, there is increasing convergence between equilibrial prices at the market and at the farmgate. As a consequence, the analysed benefit levels should become even less liable to this source of error as benefits are extrapolated to the future.

In all likelihood, the true value of benefits arising from the CGIAR is indeed much higher than any of the values presented here, as most impacts have not been assessed in a systematic and comprehensive manner. Just over half of the Centres are represented in the reviewed analyses, and even for those included, only a small number of research programmes account for most of the included benefit values. It is widely recognized that the CGIAR has achieved much more than enhanced germplasm of wheat and rice, as well as biocontrol of the cassava mealybug. The list of quantified intermediate impacts of the System alone is rather impressive, with thousands of developing-country researchers trained, thousands of studies published in peer-reviewed journals, and significant contributions made to agricultural sciences in the tropics.

Other evidence suggests that intermediate products have produced numerous substantial impacts for target poor populations in developing countries through a myriad array of complementary pathways, many of which have yet to be reliably assessed. For example, according to the "Borlaug hypothesis," improvements in agricultural productivity induced through research, which have fostered commensurate reductions in commodity prices, have reduced the profitability of production in marginal environments. In turn, such has reduced incentives for the expansion of agriculture at the margin, and has helped to avert deforestation through "land savings." Following this theory, Evenson and Rosegrant (2003) and Nelson and Maredia (1999) have shown that such impacts may run into the tens of millions of hectares of land saved from being cleared of natural biodiversity or stripped of watershed protection values. Furthermore, natural resource management research and forest policy and management research results should help to enhance such conservation impacts. However, it should also be recognized that the presence of negative environmental effects, which are potentially attributable to CGIAR research products, may mitigate some of these positive consequences (Maredia and Pingali, 2001).

To holistically assess benefits produced by a public-sector investment, the fungibility of funds invested should be considered in concert with the level of benefits produced, as the productivity of the venture does not resolve whether the private sector has been displaced. In the case of the CGIAR, the "global public goods" orientation of the System offers reason to believe that there has been mainly complementarity, rather than competition with other private and public sector research entities. Since the research products produced by the System, as with many other agricultural research organizations, are inappropriable and the beneficiaries do not have sufficient resources or the means for investment, such benefits would not haven realized without public-sector involvement. Moreover, the long time lags and high uncertainties involved in financing such research further discourage private-sector participation (Alston *et al.*, 1998b).

Furthermore, significant evidence indicates that rather than compete with existing agricultural research efforts, the existence of the IARCs catalyses other research investment (Alston *et al.*, 1998b). Evenson's (1987) econometric study offers particularly strong evidence that investment in NARS was significantly increased, so as to capture economies of scale induced through IARC activities. Thus, not only has the CGIAR impacted target populations through research outputs, but it could be credited with many of the successes of national research systems, as well, since such would have probably been smaller in the System's absence.

4.1.2 Proportion of benefits reaching target poor populations

To return to a theme mentioned in the beginning of the present study, and to put the benefits generated in a mission-relevant impact context, the proportion of benefits accruing to the poor should be considered to assess whether the CGIAR mission is being effectively fulfilled. This study cannot give a precise breakdown of benefits accruing to different groups of target beneficiaries, due to the simplicity of the analysis and the lack of relevant data in the included studies. However, significant evidence suggests that the poor have received a large portion of generated benefits, in marked contrast to some of the early critiques of the Green Revolution (GR). In many, if not most cases, adoption of improved technologies has not been disproportionately pursued by larger farmers, and in some instances, such as among rice producers in Bangladesh, smaller farmers within high-potential environments adopt innovations, including modern varieties, more frequently than do their larger counterparts (Hossain, 1998). Although it is clear that increased supply often lowers prices, and thereby may reduce revenues to nonadopters in marginal environments, in many cases these producers have benefited from increased employment opportunities in better-endowed areas, as well, through migration (Hazell and Ramasamy, 1991). Employment opportunities for these migrants are often accordingly improved through increased labour intensity, as has been repeatedly noted for Asian rice cultivation systems. Contrary to common criticism of GR technologies, the adoption of labour-saving technologies is not significantly catalysed by MV utilization, and bears greater influence from farm size and relative factor prices (David and Otsuka, 1994). Producer income increases generated through enhanced productivity also do not disappear with adopting beneficiaries, and rather filter through the rural economy, catalysing significant “multiplier effects” along the way. Consequently, it has been estimated that for each additional dollar generated in the farm sector, an additional 50 cents to one dollar is generated in the surrounding non-farm economy, as a result of increased rural demands for goods and services when farm incomes rise (Hazell and Haddad, 2001; Delgado *et al.*, 1998).

Yet, changes in producer returns only comprise a small proportion of the social gains assessed, as the bulk of analysed research benefits have been realized through prices reductions resulting from supply increases enabled via boosted productivity. For example, approximately two thirds of rice germplasm enhancement benefits in Latin America result from such price declines, while only one third accrues to producers (Sanint and Wood, 1998). For modern varieties of spring bread wheat, similar trends have been observed, notably in Pakistan (Renkow, 1993). The predominance of this impact pathway implies that the poor have received considerable shares of benefits generated through the assessed research efforts, since poorer groups spend greater proportions of their income on food (Kerr and Kolavalli, 1999). Moreover, as incomes rise, food consumption patterns often shift from those basic staple commodities which have benefited from documented widespread productivity increases attributable to the research efforts of the CGIAR, to preferred substitutes, such as meat and vegetables (Barker and Dawe, 2002; Huang and David, 1992). These low expenditure elasticities for the analysed products of CGIAR research further help to ensure that the consumer benefits derived from productivity increases accrue in substantial proportion to poorer beneficiaries. As rural-urban migration increases the number of urban poor who have few opportunities for subsistence cultivation, the importance of this impact pathway will grow. In addition, cassava mealybug biocontrol, a substantial source of estimated benefits, can be expected to principally impact poor beneficiaries, as cassava is primarily a subsistence crop in Africa, which is rarely cultivated for export purposes.

It should be noted that the reviewed assessments of surplus distribution are largely dependant on the assumption of free and undistorted markets (with the exception of Ryan, 1999), although such suppositions do not hold true for many developing countries. Moreover, it is unclear how such distortions affect the distribution of research benefits, as much hinges on how developing-country policymakers adjusted distortional measures in response to increased productivity. To the extent that the below-market procurement prices of monopsonistic parastatal "marketing boards" remained static relative to equilibrial price declines, producers received greater shares of research benefits, as these penalising policies are brought closer to liberalized market conditions (Singh, 1988; Bates, 1983). Furthermore, if productivity increases allowed for additional output to be shifted into "parallel markets" for production in excess of marketing board procurements, both consumers and producers would reap substantial benefits, as producers could receive higher quasi-market prices for greater shares of production, and the "closed" nature of these economies would allow for significant price declines. On the other hand, if procurement prices fell as quickly as equilibrial prices, and procurement quantities rose as fast as supplies increased, research benefits would accrue almost exclusively to consumers and government coffers. However, as a general trend, explicit and implicit taxation of agricultural production for food price subsidization has declined, in part due to low commodity prices on the world market, which have been resulted from increased productivity (von Braun, 1988). This pattern implies that producers have reaped a substantial share of benefits stemming from the IARCs.

Based on the evidence described in this section, which indicates that producer benefits have been relatively equitably distributed, while the balance of benefits accrues through price reductions for basic staple foods, it is plausible (and probably even conservative) to assume that impacts generated should often reach the poor in at least equal proportion to their portion of the population. Such an assumption is also consistent with analyses of the distributional consequences of untargeted food subsidies, which have found that the poor generally receive a proportional or higher allocation of benefits derived from food price reductions for commodities with low expenditure elasticities (Ahmed *et al.*, 2001; Alderman and Lindert, 1998). Consequently, if the 23% of developing country population (as of 1999) that subsists on less than one dollar per day (World Bank, 2003) are counted as the only beneficiaries, and are assumed to benefit proportionately, most scenarios still produce satisfactory results, as all "plausible" scenarios result in benefit-cost ratios over unity. Alternatively, if a further supposition ventures that the 56% of the developing world's population that presently survives on less than two dollars per day are the only beneficiaries counted, the aggregate benefit-cost ratio is more than unity in the most conservative scenario. If the more inclusive extrapolation of plausible benefits to the present is used, this measure rises to four, and if benefits accruing into the future from present research are counted, this rises further to more than eight. One or two dollars per day are crude measures of extreme absolute poverty, and many beneficiaries with incomes over these arbitrary levels are also very destitute by Western standards. If the severely resource constrained are comprehensive counted, target poor populations may be much larger than either of these somewhat crude measures indicates, and the proportion of benefits reaching poor beneficiaries may be accordingly higher. To go even further, these proportions are recent and lower than those when the CGIAR was initiated, so the "true" proportion of estimated benefits reaching target beneficiaries may be higher still. When these results are combined with the fact that most benefits of the CGIAR probably remain unestimated, there is a clear indication that the CGIAR investment has been efficient and effective. It remains as a future challenge for impact assessment to more comprehensively represent these benefits.

4.2 Methods of economic impact assessments reviewed

During the review process it became apparent that economic assessments of impact within the System could benefit from enhancement of scope and methods applied. All of the studies reviewed could be regarded as financial, rather than economic analyses, as market prices are not adjusted so as to include external effects or to compensate for market distortions (through the application of social or shadow prices). Rather, the studies make the implicit assumption that price equals marginal values, even when such fluctuate drastically due to external influences. Many commodities' markets are known to be highly distorted, due to external and internal government policies, so a social, rather than financial orientation may be much more appropriate and illustrative.

The fact that all of the reviewed studies take a financial, rather than social, approach runs counter to the recommendations of Alston *et al.* (1996) that "we ought to take into account the total effects on the welfare of all affected groups when we can." When a structure for impact assessment at the System level was being initially formulated, a "need to cover both the intended and unintended effects of research at different stages from various forms of uptake of research results to their ultimate impact on target groups or objects (e.g., farmers, environment, society, economy, etc.)" was noted by the Impact Assessment Task Force (Özgediz, 1995). Similarly, since the 1970s, the World Bank has recommended application of the following techniques to facilitate social, rather than financial, analysis (Mosley, 2001):

- Sectoral conversion factors
- Shadow prices for inputs and products
- Tracing of generated income
- Distributional weights for income generation
- Conversion factors for non-traded goods

None of these techniques has been fully applied within the reviewed analyses. To be fair, most multilateral agencies have failed to effectively apply these methodologies to the majority of benefit-cost studies commissioned. However, this failure does not eliminate the necessity of including such factors. While it may be exceedingly difficult to operationalize the inclusion of external costs, and utilize social pricing, little methodological progress can be made if such is not attempted. Simply ignoring these factors does not render them irrelevant, and can create somewhat misleading results, which reduce assessment credibility. For example, many of the most successful CGIAR innovations catalysed increases in input use (such as the initial introduction of semi-dwarf varieties of wheat and rice, which do not lodge under high doses of fertilizer), and the social costs of changes in input-use associated with adoption should be considered in a more meaningful manner. However, it should be noted that some progress has been made in assessing these external effects in separate small-scale assessments, although additional efforts will be needed to integrate findings into large-scale analyses (Pingali, 2001).

Unfortunately, few of the studies utilized an explicit counterfactual scenario. Those that do analyse prevention of losses due to pests or diseases, which is a type of benefit that makes the need for counterfactual scenarios particularly apparent. Virtually all major textbooks on economic project appraisal place a great deal of emphasis on counterfactual development, so it is somewhat surprising that such is attempted so rarely in IARC assessments. Although overt counterfactual postulates may seem somewhat arbitrary and contrived, implicit assumptions for the "without" scenario are even more so. Counterfactual plausibility should also be enhanced through an empirically-derived basis, such as through the identification of control populations. Every impact assessment contains a counterfactual, whether explicit or implicit, and, thus, methodological transparency demands that the assumptions behind such be overtly declared.

The lack of a counterfactual also may have contributed to the meagre consideration of mitigating influences apparent in the studies. While researchers may have an *a priori* understanding that the research outputs studied have been the prime causal factors for the impact trends observed, this causality is not so readily apparent to the "outside observer." Arguably, in science it is much easier to disprove than to prove, so refuting the causality of other relevant factors would be quite effective for establishing causality. The somewhat simple approach of the reviewed studies leaves them vulnerable to the impact claims of alternative complementary interventions. For example, the findings of Chavas (2001) that essentially all yield growth may be explained by changes in input use, and that there has been no overall technological development in the agricultural sector are more difficult to refute if input use remains superficially considered in the bulk of assessments.

Since so few of the included studies attempt institutional attribution, potentially controversial conflict-prone claims to research impact are avoided. However, this also serves to undermine the accountability purpose of the IAs, as claims of collective credit do not necessarily offer strong arguments for allocation to individual research entities. Thus, some sort of universally acceptable means of including alternative research providers in counterfactuals, so as to partition impacts is necessary, in order to improve the potential impact of IA on allocative processes.

While these somewhat simple benefit-cost techniques, as presently applied, are valuable for demonstration of investment productivity, such do not have high potential for reliably feeding into priority-setting processes. This limited capability is largely a product of the simplicity of the analytical techniques commonly utilized, along with the paucity of reliable data available to allow the derivation of precise intermediate impact estimates. In the absence of high-quality, comprehensive empirical data on control groups, productivity, crop management, or input use, analysts have to make due with only very limited sets of information, from which only crude extrapolations may be made. With only rough results generated, few recommendations for priority setting can be reliably extrapolated. Since donors have been keen to request "lessons learned," which are relevant to current programmatic implementation, more comprehensive and representative data sets will be necessary to dependably generate such lessons. To foster the collection of such data, additional resources will need to be devoted to IA within the System.

At an even more basic level, many of the reviewed studies appear to be characterized by a serious lack of methodological transparency, which undermines the accountability objectives inherent in the pursuit of economic impact assessment. Data collection methods employed in the reviewed studies are not frequently explained or cited. For example, key meta-data, such as survey sample sizes, are omitted, while major sources of information are often not explicitly mentioned. The treatment of data is also often inadequately described, so that discount rates, commodity prices and the aggregation of yield increase estimates are not specified. In certain cases such details are described in referenced unpublished internal publications, but such are not easily available to the casual reader. Such opacity reduces the reliability of derived estimates and may make results inadequate for establishing credible impact claims, since the basis for such is unclear. If impact assessors wonder "why has impact assessment not made more of a difference?," such a lack of clarity is likely to be one viable answer.

In general, ensuring the quality of basic data utilized, rather than focusing principally upon data treatment, should become more of a focus for economic impact assessment efforts. Operationalising this suggestion would require significant changes to the manner in which impact assessments are pursued within the System. Since IA is pursued as research endeavour, rather than as a procedural requirement, the economists who conduct such studies have little incentive to undertake exhaustive and time-consuming data collection efforts, and instead focus on methodological innovation. The lack of a clearly defined role for IA does nothing to aid this situation, as the standards necessary for fulfilling different envisioned roles have never been identified, and allocative decisions do not appear to be tied to the quality with which impacts are analysed. As such, and somewhat ironically, the marginal value of investing resources in impact assessment work has not been proven, as compared with other social science research options. Consequently, research managers have little reason to prioritize impact assessment work, and the effectiveness of the CGIAR investment is not comprehensively represented, nor can "lessons" be effectively "learnt" to enhance future efficacy.

4.3 Recommendations for future activities

To move beyond the wide range of plausible estimates developed in the present analysis, and "zero in" on a more precise benefit range, a greater degree of consensus needs to be established for expectations from ex-post impact assessment. Assessing benefits with a fixed amount of resources generally presents a tradeoff between scope and reliability, and client preferences are needed to select an optimal allocation between these two attributes. To achieve consensus on such allocation, it is critical that the clients of IA research articulate minimum standards for impact claims. Such standards should not be established by SPIA, Centres or single experts in isolation, as the diverse views of investors and other stakeholders will not be comprehensively represented through such a proxy. It is also critical that dialogue be established among impact assessors and intended audiences, so that such expectations are effectively communicated, and such can realistically incorporate the constraints imposed by limited resources and competing priorities.

The IA methodological literature is replete with "best practice" manuals and guiding principles (Alston *et al.*, 1996; Baker, 2000; Echeverría, 1990; Maredia et al, 2000) but a clear gulf persists between these idealized conceptions and operationalized assessments. Repeatedly, it is stressed that techniques, such as shadow prices be utilized to encompass environmental effects, that conversion factors should help to estimate distributional impacts, and that counterfactuals should be explicit and empirically-derived. Yet, as has been noted in the present document, such techniques are nearly never incorporated into large-scale assessments of research impact, and the simplistic techniques typically applied are rarely criticized for these omissions.

Simultaneously, documentation of impact is increasingly emphasized as a necessary requirement for programmatic accountability. Furthermore, in the methodological literature, impact assessment is often assumed to be a primary source of information for priority-setting procedures (Alston *et al.*, 1996). However, there is no clear and consistent linkage between documented impact and resource allocation. Ironically, those research programmes for which impact has been most thoroughly documented (commodity breeding) have suffered declining budgets, while those for which there have been very few large-scale impacts assessed (natural resource management) have benefited from rising allocations. This dichotomy clearly presents a potential quandary for the accountability role of impact assessment.

The apparent low ability of impact assessments to influence allocative decisions may be principally caused by either the possible low persuasiveness of the current impact assessment portfolio or by overriding concerns unrelated to past performance, which may dominate funding processes. Since the latter has not been analysed in this study, the former is focussed upon in the activities recommended.

The criteria enumerated in the present meta-analysis for critically reviewing the included studies, and the range of benefits presented under the different scenarios may constitute a viable context for initiating improved stakeholder-assessor dialogue. It is recognized that this would not be the first attempt to gain better understanding of client needs, but it is possible that prior attempts may have had few meaningful results due to the absence of a specific context for eliciting responses. It is difficult to abstractly define minimum general data standards, when impacts are so diverse in nature, and the methodological approaches lack an accepted norm. However, if a response to a specific study is requested, it is much more likely than meaningful feedback may be provided. Thus, if a selection of studies encompassing a variety of degrees of entailed effort are to be reviewed by a panel of investor representatives, with the intention of producing a consensus evaluation, trends in expectations may become evident. To ensure that such a workshop is focussed on critical review, rather than synopsis, it would be beneficial for donors to both present and critique Centre assessments. Thus, each investor/stakeholder could bear responsibility for a short summary of content, strengths, weaknesses and points for improvement of a specific IARC impact study. In turn, each presentation would be followed by discussion intended to produce a consensus impression of study quality.

Patterns in expectations evident in the workshop could then be distilled into minimum IA standards, which are broadly acceptable to IA audiences. With these standards established, the studies reviewed in the meta-analysis may be revisited. To facilitate more comprehensive impact coverage in this second attempt at aggregation, the IARCs should be invited to submit additional and/or revised impact studies for inclusion. Once submission has taken place, review should centre on the client-derived standards, and evaluate to degree to which such are met by the reviewed assessments. The resulting aggregate benefit values should then be implicitly acceptable to target IA audiences, and insights from the critical review could be reliably utilized to identify needs for IA improvement.

5. CONCLUSION

In summation, based on the results of the present analysis, five key conclusions may be drawn:

- The productivity of the CGIAR investment has been shown in all five scenarios of aggregate benefits estimates to exceed normal standards of investment efficacy by a substantial margin.

- Three main categories of innovations constitute the vast majority of plausibly estimated benefits: modern varieties of wheat, modern varieties of rice and the biological control of the cassava mealybug.
- Selection of a single scenario as the most “true” minimum aggregate benefit-cost ratio of the CGIAR investment is arbitrary without client articulation of required evidence for impact claims.
- More comprehensive data collection procedures, including a stakeholder-needs-based mechanism for “priority-setting for impact-assessment,” could help to render more reliable and meaningful results from IA studies.
- Improving the average degree of methodological transparency would be an easy means to raise the confidence with which findings may be utilized and scaled-up in future analyses.

The effectiveness and efficiency of CGIAR research could be more substantially demonstrated through improved impact assessment procedures. In general, the realm of impact assessment within the CGIAR is in need of additional guidance at the System level, since methodological standards for impact assessment vary so greatly among Centres and programmes. Unfortunately, institutional dynamics have inhibited the development of this direction, as it appears that insufficient resources have been allocated to System level guidance.

With the range of potential benefit-cost ratios for the CGIAR investment ranging from less than two to over 17, it is logical to ask: What ratio should be considered as closest to the “truth,” and should be used for future reference? This study cannot provide a definitive answer to this question, as such is dependent upon the standards of causality demanded from IA for the acceptance of results. In turn, specification of such standards must come from the “clients” of IA studies, principally investors in the System, so that the kinds of proof required for the validity of impact claims can be identified.

It is crucial that the necessary data requirements for staking different kinds of impact claims be articulated by key stakeholders. Such would help to clarify the resources to be invested in IA activities, and would help to improve IA utility for priority-setting and allocative decisions. The foundation for such requirements should be interaction with the prospective “clients” of impact assessment studies, including donors, co-sponsors, NARS, NGOs, farmers’ organizations and even researchers within the CGIAR institutions, so as to achieve consensus on the minimum requirements for viable impact assertion, and the proper role of impact assessment. From this basis, a System level impact-assessment prioritization system should be derived, as was initially recommended by the Impact Assessment Task Force eight years ago. Once such a system has been developed, impact assessment for accountability should be planned at a System level in consultation and consensus with the IARCs.

To conclude, much credit should go to those IARCs that have successfully initiated endogenous economic impact assessment efforts, with little System-level guidance and limited resources. These efforts, even if somewhat uncoordinated, have helped to facilitate methodological progress, and have helped to highlight System successes. Now, to move forward, and help build upon these accomplishments, impact assessors and interested stakeholders should come together to define expectations and standards for the future.

6. REFERENCES

Ahmed, A.U., H.E. Bouis, T. Gutner & H. Löfgren. 2001. The Egyptian Food Subsidy System: Structure, Performance, and Options for Reform. *Research Report No. 119*. Washington, DC: IFPRI.

Alderman, H. & K. Lindert. 1998. The Potential and Limitations of Self-Targeted Food Subsidies. *World Bank Research Observer*. 13(2): 213-229.

Alston, J.M., C. Chan-Kang, M.C. Marra, P.G. Pardey & T.J. Wyatt. 2000. A Meta-Analysis of Rates of Return to Agricultural R& D: Ex Pede Herculem? *Research Report 113*. Washington, DC: IFPRI.

Alston, J. M., B. J. Craig & P. G. Pardey. 1998a. Dynamics in the creation and depreciation of knowledge, and the returns to agricultural research. *EPTD Discussion Paper No. 35*. Washington, DC: International Food Policy Research Institute.

Alston, J.M., G.W. Norton & P.G. Pardey. 1996. *Science under scarcity: Principles and practice for agricultural research evaluation and priority setting*. Ithaca, NY: Cornell University Press.

Alston, J.M., P.G. Pardey & J. Roseboom. 1998b. Financing agricultural research: International investment patterns and policy perspectives. *World Development*. 26(6): 1057-1071.

Anderson, J.R. 1998. Selected Policy Issues in International Agricultural Research: On Striving for International Public Goods in an Era of Donor Fatigue. *World Development*. 26(6): 1149- 1162.

Anderson, J.R. 1997. *On Grappling with the Impact of Agricultural Research*. Prepared for International Centres Week, Washington, DC. Unpublished manuscript.

Baker, J. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington, DC: World Bank.

Bantilan, M.C.S. & P.K. Joshi. 1996. Returns to Research and Diffusion Investments on Wilt Resistance in Pigeonpea. *Impact Series no. 1*. Andhra Pradesh, India: ICRISAT.

Barker, R. & D. Dawe. 2002. The transformation of the Asian rice economy and directions for future research: the need for increased productivity. *In* M. Sombilla, M. Hossain and B. Hardy (eds.), *Developments in the Asian Rice Economy*. Los Banos, Philippines: IRRI.

Bates, R.H. 1983. Patterns of Market Intervention in Agrarian Africa. *Food Policy*. 8(4): 297-304.

Binswanger, H.P. & K. Deininger. 1997. Explaining Agricultural and Agrarian Policies in Developing Countries. *Journal of Economic Literature*. 35(4): 1958-2005.

Bokonon-Ganta, A.H., H. de Groote & P. Neuenschwander. 2002. Socio-economic impact of biological control of mango mealybug in Benin. *Agriculture, Ecosystems and Environment*. 93: 367-378.

Byerlee, D. & P. Moya. 1993. *Impacts of International Wheat Breeding Research in the Developing World, 1966-90*. Mexico, D.F.:CIMMYT.

Byerlee, D. & G. Traxler. 1995. National and International Wheat Improvement Research in the Post-Green Revolution Period: Evolution and Impact. *American Journal of Agricultural Economics* 77: 268-278.

Chavas, J.P. 2001. An International Analysis of Agricultural Productivity, *In* L. Zepeda, (ed.) *Agricultural Investment, Production Capacity, and Productivity*. Rome: FAO [downloadable from http://www.fao.org/docrep/003/x9447e/x9447e00.htm]

Cooksy, L. 1997. *Methodological Review and Synthesis of Existing Ex Post Impact Assessments*. Rome: Impact Assessment and Evaluation Group, CGIAR.

Dalton, T.J. & R.G. Guei. 2003. Ecological Diversity and Rice Varietal Improvement in West Africa. *In* R.E. Evenson and D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

David, C.C. & K. Otsuka (eds). 1994. *Modern rice technology and income distribution in Asia*. Manila, Philippines: IRRI.

Delgado, C., J. Hopkins & V.A. Kelly. 1998. Agricultural Growth Linkages in Sub-Saharan Africa. *IFPRI Research Report 107*. Washington, DC: IFPRI.

Echeverría, R.G. (1990). Assessing the impact of agricultural research. *In*: R.G. Echeverría (Ed.). *Methods for diagnosing research system constraints and assessing the Impact of agricultural research* (Vol. 2). The Hague: International Service for National Agricultural Research (ISNAR).

Evenson, R.E. 1987. The International Agricultural Research Centers: Their impact on spending for national agricultural research and extension. *CGIAR Study Paper No. 22*. Washington, DC: CGIAR Secretariat.

Evenson, R.E. & D. Gollin. 1997. Genetic resources, international organizations and improvement in rice varieties. *Economic Development and Cultural Change*. 45(3): 471-500.

Evenson, R.E. & M. Rosegrant 2003. The Economic Consequences of CGI Programs. *In* R.E. Evenson and D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Fuglie, K.O., L. Zhang, L.F. Salazar & T.S. Walker. 1999. *Economic Impact of Virus-Free Sweetpotato Planting Material in Shandong Province, China*. Lima, Peru: International Potato Center.

Griliches, Z. 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica*. 25(4): 501-522.

Hassan, A. Aw, K. Shideed, S. Ceccarelli, W. Erskine, S. Grando & R. Tutwiler. 2003. The Impact of International and National Investment in Barley Germplasm Improvement in the Developing Countries. *In* R.E. Evenson & D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Hazell, P. & L. Hadadd. 2001. *Agricultural Research and Poverty Reduction*. Discussion Paper No. 34. Washington, DC: IFPRI.

Hazell, P.B.R. & C. Ramasamy (eds). 1991. *The Green Revolution reconsidered: the impact of high-yielding rice varieties in South India*. Washington, DC: IFPRI.

Heisey, P.W., M.A. Lantican & H.J. Dubin. 2002. *Impacts of International Wheat Breeding Research in Developing Countries, 1966-97*. Mexico, D.F.: CIMMYT.

Heisey, P.W., M.A. Lantican & H.J. Dubin. 2003. Wheat. *In* R.E. Evenson and D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Hossain, M., D. Gollin, V. Cabanilla, E. Cabrera, N. Johnson, G.S. Khush & G. McLaren. 2003. International Research and Genetic Improvement in Rice: Evidence from Asia and Latin America. *In* R.E. Evenson and D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Hossain, M. 1998. Rice research, technical progress, and the impact on the rural economy: the Bangladesh case. In Pingali, P.L. & M. Hossain (eds.). *Impact of Rice Research*. Manila, Philippines: IRRI.

Huang, J. & C.C. David. 1992. Demand for cereal grains in Asia: the effect of urbanization. *Agricultural Economics*. 8: 107-124.

Johnson, N.L., D. Pachico & C.S. Wortmann. 2003. The Impact of CIAT's Genetic Improvement Research on Beans. *In* R.E. Evenson & D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Johnson, N.L. & V.M. Manyong. 2003. The Impact of IARC Genetic Improvement Programs on Cassava. *In* R.E. Evenson & D. Gollin (eds.). *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research.* Oxon, U.K.: CABI.

Kerr, J. & S. Kolavalli. 1999. *Impact of agricultural research on poverty alleviation: conceptual framework with illustrations from the literature*. Environment and Production Technology Division Discussion Paper No. 56. Washington, DC: IFPRI.

Maredia, M. & P. Pingali, 2001. *Environmental impacts of productivity-enhancing crop research: a critical review*. Rome, Italy: CGIAR Technical Advisory Committee Secretariat, FAO.

Maredia, M., D. Byerlee & J.R. Anderson. 2000. Ex Post Evaluation of Economic Impacts of Agricultural Research Programs: A Tour of Good Practice. *In The Future of Impact*

Assessment in the CGIAR: Needs, Constraints and Options. Rome, Italy: CGIAR Technical Advisory Committee Secretariat, FAO.

Marasas, C.N., M. Smale & R.P. Singh, 2002. *The impact of agricultural maintenance research: the case of leaf rust resistance breeding in CIMMYT-related spring bread wheat*. Mexico, D.F.: CIMMYT.

Morris, M.L. 2002. *Impacts of International Maize Breeding Research in Developing Countries, 1966-98*. Mexico, D.F.: CIMMYT.

Morris, M., M. Mekuria & R.Gerpacio. 2003. Impacts of CIMMYT Maize Breeding Research. *In* R.E. Evenson. *Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research*. Oxon, U.K.: CABI.

Mosley, P. 2001. A simple technology for poverty-oriented project assessment. *Impact Assessment and Project Appraisal*. 19(1): 53-67.

Nelson, M. & M. Maredia. 1999. *Environmental Impacts of the CGIAR – An Initial Assessment*. Rome, Italy: Technical Advisory Committee Secretariat, FAO.

Nin, A., C. Arndt & P.V. Preckel. 2003. Is agricultural productivity in developing countries really shrinking? New evidence using a modified nonparametric approach. *Journal of Development Economics*. 71: (2003) 395-415.

Norgaard, R.B. 1988. The Biological Control of Cassava Mealybug in Africa. *American Journal of Agricultural Economics*. 70: 366-371.

Oehmke, J.F. 1988. The Calculation of Returns to Research in Distorted Markets. *Agricultural Economics*. 2: 291-302.

Özgediz, S. 1995. Strengthening *Evaluation in the CGIAR: Needs and Options*. CGIAR Task Force on Impact Assessment. Washington, DC: CGIAR Secretariat.

Pingali, P.L. 2001. *Milestones in Impact Assessment Research in the CGIAR, 1970-1999*. Mexico, D.F.: CGIAR Standing Panel on Impact Assessment.

Renkow, M. 1993. Differential technology adoption and income distribution in Pakistan: Implications for research resource allocation. *American Journal of Agricultural Economics*. 75(1): 33-43.

Reynolds, M., S. Rajaram & K.D. Sayre. 1999. Physiological and genetic changes of irrigated wheat in the post-Green Revolution period and approaches for meeting projected global demand. Crop *Science.* 39(6): 1611-1621.

Ryan, J.G. 1999. Assessing the Impact of Rice Policy Changes in Viet Nam and the Contribution of Policy Research. *Impact Assessment Discussion Paper No. 8*. Washington, DC: IFPRI.

Sah, R.J. & T.N. Srinivasan. 1987. Distributional Consequences of Rural Food Levy and Subsidized Urban Rations. *European Economic Review*. 32: 141-159.

Sanint, L.R. & S. Wood. 1998. Impact of Rice Research in Latin America and the Caribbean During the Past Three Decades. In Pingali, P.L. and M. Hossain (eds.) *Impact of Rice Research*. Manila, Philippines: IRRI.

Sayre, K.D., S. Rajaram & R.A. Fischer. 1997. Yield potential progress in short bread wheats in northwest Mexico. *Crop Science*. 37(1): 36-42.

Shultz, T.W. 1964. *Transforming Traditional Agriculture*. New Haven: Yale University Press.

Singh, H.V. 1988. An Extension of the Simple Analytics of Segmented Grain Markets and the Case for Liberalization. *World Development*. 16(6): 759-763.

TAC Secretariat, 2001. The future of impact assessment in the CGIAR - Needs, constraints, and options: Proceedings of a workshop organized by the Standing Panel on Impact Assessment of the Technical Advisory Committee 3-5 May 2000, FAO, Rome. Rome, Italy: FAO.

von Braun, J. 1988. Implications of Consumer-Oriented Food Subsidies for Domestic Agriculture. *In*. Pinstrup-Andersen, P. (Ed.) *Food subsidies in developing countries: costs, benefits, and policy options*. Baltimore, MD, USA: Johns Hopkins University.

World Bank. 2003. Millennium Development Goals: Eradicate Extreme Poverty. [accessed from http://www.developmentgoals.org/Poverty.htm#percapita on 20/3/2003].

Zeddies, J., R.P. Schaab, P. Neuenschwander & H.R. Herren. 2001. Economics of biological control of cassava mealybug in Africa. *Agricultural Economics*. 24: 209-219.

FIELDS FOR STUDY REVIEW DATABASE

Study Title:

Author:	Author Institution(s):
Citation:	Year of Study:
Geographic Coverage:	Description of Innovation:
Overall Advantage of Innovation:	Indicator of Productivity Changes:
Measurement Method for Productivity Changes:	Diaggregation of Productivity Changes:
Measurement Method for Utilization/Adoption:	Changes in Inputs, Crop Management Considered:
Mitigating Factors Considered:	Basis of Credit Partitioning:
Other Actors Attributed:	Basis of Counterfactual:
Research Partners Included in Counterfactual:	Disaggregation of Counterfactual:
Basis of Input Prices:	Basis of Commodity Prices:
Elasticity of Supply:	Elasticity of Demand:
Assummed Lag Period (yrs):	Start Year of Benefits:
End Year of Benefits:	Base Year of Study:
Factors Tested in Sensitivity Analysis:	Total NPV Under Main Assumptions:
Low Range of Total NPV:	Most Credible NPV Estimate:
High Range of Total NPV:	Main Peak Annual Benefits:
Low Peak Ann Benefits:	Comments on Analysis:
Assumptions for Most Credible NPV Estimate:	High Peak Annual Benefits:
CGIAR Role in Development of Innovation:	Main IARC NPV:

APPENDIX II

Characteristics of studies producing benefit values included in the meta-analysis

Citation	Type of Innovation	Type of Analysis	Geographic Coverage	Period Assessed	Measurement Method for Utilisation/Adoption	Measurement Method for Productivity Changes	Basis of Credit Partitioning	Basis of Commodity Prices	Adjusted "Plausible" Benefits (billions of 1990 $)	Rating
Bantilan, M.C.S. and P.K. Joshi. 1996. Returns to Research and Diffusion Investments on Wilt Resistance in Pigeonpea. Impact Series no. 1. Andhra Pradesh, India: ICRISAT.	Modern varieities of pigeonpea	Economic surplus with disaggregation by state	India	1975 - 2005	On farm surveys were conducted to measure adoption production gains, as well as changes in production costs. The survey samples were selected so as to be representative of two adoption regimes - the target zone and the farmer to farmer diffusion zone. A reconnaissance survey of pigeonpea growing districts and supplementary secondary data were used to identify survey sites of high pigeonpea importance, while stratified multi-stage random sampling of pigeonpea growing areas was used to identify specific farms for surveys.	The same on farm surveys as for adoption were used to generate data for analysis of changes in unit production costs.	No attribution in study	Domestic prices were estimated through the on-surveys for adoption and unit cost changes	0.22	Significantly demonstrated
Bokonon-Ganta, A.H., H. de Groote, and P. Neuenschwander. 2002. Socio-economic impact of biological control of mango mealybug in Benin. Agriculture, Ecosystems and Environment. 93: 367-378.	Biocontrol of mango mealybug	Economic surplus, disaggregated by department	Benin	1987 - 2011	Adoption not necessary, as biocontrol agent spreads naturally	A total of 300 producers were subjected to three semi-structured interviews between 1989 and 1991, as well as a final survey of the producers' trees in 1999. Of these, 142 sets of responses were considered reliable, and were used for estimating production levels over time, as well as perceptions and knowledge of the biocontrol programme's impact.	No attribution in study	Domestic prices (basis not specified)	0.11	Plausible
Byerlee, D., and G. Traxler. 1995. National and International Wheat Improvement Research in the Post-Green Revolution Period: Evolution and Impact. American Journal of Agricultural Economics 77: 268-278.	Modern varieties of Spring bread wheat	Economic surplus with disaggregation into 4 "mega-environments" and 4 regions	Global	1966 - 1990	Adoption estimates were based on Dalrymple's 1977 survey and the 1990 CIMMYT Survey of Wheat Research Impacts. "Area estimates were based on annual government surveys in some countries, special surveys at a regional or country level, seed sales in some countries, and wheat researchers estimates," but it is not explained how often each technique was applied.	Yield gain was estimated based on rough extrapolation from relative gains experienced at experiment stations, and was broken down into Stage 1 (traditional to modern variety) and Stage 2 (modern variety to newer modern variety) gains. Stage 1 gains were estimated as cumulative percentages, while Stage 2 gains were estimated as annual increments.	The percentage CIMMYT credit figure mentioned is an average of the cumulative 1990 values. The percentage of germplasm with CIMMYT origin is used as a basis for partitioning, with 85% CIMMYT credit for CIMMYT crosses, and 50% for NARS' crosses with CIMMYT parent.	Rotterdam CIF import price for net importers, average of import price and FOB Gulf Ports price for exporters, deflated by PPI to 1990 base	9.75	Significantly demonstrated
Dalton, T.J. & R.G. Guei. 2003. Ecological Diversity and Rice Varietal Improvement in West Africa. In R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Modern varieties of rice	Monte-Carlo Simulation	West Africa	1970 - 1998 (benefits only calculated for 1998)	Adoption values were based on expert opinion estimates by national research and extension authorities, which were used as upper limits in Monte Carlo modelling of uncertain parameters. These values were applied to FAO statistics for national rice areas of cultivation.	Not clearly described	Attribution is based on the proportion of CGIAR derived germplasm in the adopted varieties.	Domestic FAO-sourced import parity prices	0.15	Plausible
Fuglie, K.O., L. Zhang, L.F. Salazar, and T.S. Walker. 1999. Economic Impact of Virus-Free Sweetpotato Planting Material in Shandong Province, China. Lima, Peru: International Potato Center.	Seed technology	Economic surplus	Shandong Province, China	1988 - 2020	Group interviews with farmers in 30 villages, were used to estimate adoption, in combination with seed production data from the provincial seed programme	The same group interviews with farmers in 30 villages as for adoption were used to estimate productivity enhancement, and this was validated against demonstration plots	The main technique is entirely attributed to CIP, but was applied locally by Chinese institutions. However, the study notes that this technique would have probably reached the province eventually, even without CIP's participation. Thus, when an IARC NPV is presented below, it is based on the technique being applied 5 years later in CIP's absence.	Domestic prices from interviews	0.25	Significantly demonstrated
Hassan, A. Aw, K. Shideed, S. Ceccarelli, W. Erskine, S. Grando and R. Tutwiler. 2003. The Impact of International and National Investment in Barley Germplasm Improvement in the Developing Countries. In R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Modern varieties of barley	Economic surplus with national disaggregation	Global	1980 - 1997	Based on NARS data, average adoption of improved barley cultivars is estimated for 1980-1997.	A combination of on-farm trials and experiment station data are utilised for deriving the k-factor for productivity increases (these are cited from unpublished studies, and no metadata are provided)	No attribution in study	Average domestic prices from the FAO Trade Yearbook	0.33	Plausible
Heisey, P.W., M.A. Lantican, and H.J. Dubin. 2002. Impacts of International Wheat Breeding Research in Developing Countries, 1966-97. Mexico, D.F.:CIMMYT.	Modern varieties of wheat	Economic surplus	Global	1966 - 1997 (benefits only calculated for single year in late 1990s)	Adoption data come from an update of Byerlee and Traxler's work, and the estimates are derived from an extensive survey of public breeding institutes and seed companies.	Based on values reported in a few other macro-level studies, genetic improvement based yield gain is estimated to average between 0.2 and 0.4 tonnes per hectare.	Attributive coefficients are based on the average proportion of CIMMYT derived germplasm in adopted modern varieties.	World market average price	0.88	Plausible
Hossain, M., D. Gollin, V. Cabanilla, E. Cabrera, N. Johnson, G.S. Khush, and G. McLaren. 2003. International Research and Genetic Improvement in Rice: Evidence from Asia and Latin America. In R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Modern varieties of rice	Economic surplus with national disaggregation	Asia (Latin America is discussed, but not in terms of economic impacts)	1965-1999 (benefits only calculated for single year in late 1990s)	National adoption estimates are sourced from the World Rice Statistics Database (no metadata concerning the database are provided in the study)	Costs and return data have been collected by NARS and IRRI during the late 1990s through household surveys (no metadata or citations are provided for the surveys)	No attribution in study	Domestic prices	4.31	Plausible
Hossian, M. 1998. Rice research, technical progress, and the impact on the rural economy: the Bangladesh case. In Pingali, P.L. and M. Hossain (eds.) Impact of Rice Research. Manila, Philippines: International Rice Research Instiute.	Modern varieties of rice	Economic surplus	Bangladesh	1973 - 1993	Review of 3 published studies and unpublished survey data from national research institutes of Bangladesh (no metadata provided on data sources for cited studies)	Same sources as for adoption - review of three published studies and unpublished survey data from national research institutes of Bangladesh	No attribution in study	Domestic prices for unit cost reductions	3.21	Significantly demonstrated
Johnson, N.L. and V.M. Manyong. 2003. The Impact of IARC Genetic Improvement Programs on Cassava. In R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Modern varieties of cassava	Economic surplus with national disaggregation	Global	1970 - 1998	FAO area estimates are used to determine the total areas under cassava, while LAC and Asian adoption information comes primarily from estimates made by Clair Hershey, cassava breeder, in consultation with other cassava scientists. Information from annual reports and other internal documents was also included. In an internal publication by Hershey adoption estimates by country are provided for 1988, as well as diffusion curves by region. For Africa, the area under improved varieties is multiplied by IARC percentage of varieties released to obtain the area under IARC varieties.	Yield gain information comes primarily from estimates made by Clair Hershey, cassava breeder, in consultation with other cassava scientists, complemented by internal publications. Since the yield gains utilised are cited from experiment station data, the estimates are decreased by 40% to account for lower on-farm gains.	No attribution in study	Domestic root prices were used for Asia (FAO), while the world price was used for starch, since national level starch prices were unavailable. For Latin America, a world root price estimated by Hershey was used since national level data are not available.	0.23	Plausible (for Asia and Latin America)

Johnson, N.L., D. Pachico, and C.S. Wortmann. 2003. The Impact of CIAT's Genetic Improvement Research on Beans. In R.E. Evenson and D. Gollin (eds.). Crop Variety Improvement and its Effect on Productivity: The Impact of International Agricultural Research. Oxon, U.K.: CABI.	Modern varieties of beans	Economic surplus with national disaggregation	Global	1979 - 1998	Most adoption data for Latin America are cited from empirical adoption studies published as internal working papers, while estimates for Brazil are based on seed sales data and expert opinion. African adoption proportions are based on internal empirical studies and estimates of the Pan African Bean Research Alliance. Adoption estimates for both regions are multiplied by FAO statistics of total bean areas planted in each country to derive area estimates for adoption.	Most national yield gain estimates for Latin America were based on the review of internal adoption studies used for estimating national areas of adoption, and these use farm estimates for yield advantages (no metadata are provided). The African gain estimates are derived from an unpublished study of the Pan African Bean alliance (no metadata are provided).	No attribution in study	National prices from FAO where available, prices from neighbouring countries if national prices were not obtainable	0.59	Plausible
Morris, M.L. 2002. Impacts of International Maize Breeding Research in Developing Countries, 1966-98. Mexico, D.F.:CIMMYT.	Modern varieties of maize	Economic surplus	Global	1966 - 1998 (benefits only calculated for single year in late 1990s)	Global assessment of the adoption of CIMMYT-derived germplasm, based on the results of a worldwide survey of 267 private seed companies and 104 public maize breeding institutes.	Yield gains due to the adoption of CIMMYT-derived varieties are estimated to range between 15% and 45%, based on experiment station yield data suggesting that there has been a 1 - 2% annual gain in genetic potential during the study period of 1966-98. Four yield gains estimates within this range are multiplied by an average MV maize yield of 3.5 tonnes/ha to estimate productivity effects, and the values produced are halved to account for complementary factors .	To attribute CIMMYT, the resulting net benefit value is multiplied by 25%, 50%, or 75%.	World market average price	0.44	Plausible
Ryan, J.G. 1999. Assessing the Impact of Rice Policy Changes in Viet Nam and the Contribution of Policy Research. Impact Assessment Discussion Paper No. 8. Washington, DC: IFPRI.	Technical assisstance to rice market liberalisation	Application of Viet Nam Agricultural Spatial Equilibrium model	Vietnam	1995 - 2000	A total of 38 interviews with government personnel were used to determine the degree to which these changes in Vietnamese rice policies were affected by IFPRI's recommendations, and there was a general consensus that IFPRI accelerated the pace of these changes. Policy changes were also analysed to assess the degree to which specific IFPRI recommendations were embedded.	The Viet Nam Agricultural Spatial Equilibrium model, developed as part of IFPRI's technical assistance package for informing the evolution of the Vietnamese rice policy, was applied to determine the benefits derived through the observed policy changes.	Implementation of recommended policies one year earlier than would be the case without IFPRI's assistance.	World market prices for exports, domestic prices for domestic consumption	0.04	Significantly demonstrated
Sanint, L.R. and S. Wood. 1998. Impact of Rice Research in Latin America and the Caribbean During the Past Three Decades. In Pingali, P.L. and M. Hossain (eds.). Impact of Rice Research. Manila, Philippines: International Rice Research Instiute.	Modern varieties of rice	Application of DREAM multimarket economic-surplus model	Latin America and the Caribbean	1967 - 1995	Review of 10 previously published regional rice studies for Latin America (no metadata provided on data sources for cited studies)	Basic productivity data were derived from the 10 regional studies used for adoption estimates, of which exogenous sources of yield growth are independently specified, based on the same data set.	No attribution in study	Domestic prices for unit cost reductions	8.28	Significantly demonstrated
Zeddies, J., R.P. Schaab, P. Neuenschwander, and H.R. Herren. 2001. Economics of biological control of cassava mealybug in Africa. Agricultural Economics. 24: 209-219.	Biocontrol of cassava mealybug	Economic surplus with disaggregation into 3 "ecological zones"	Africa	1974 -2013	Adoption not a factor, as control agent (parasitic wasp) dispersed naturally, estimates of proliferation and spread based on entomological studies	Crop loss reductions to stable equilibria (return to pre-invasion productivity levels) for the 3 "ecological zones," based on 6 entomological studies; baseline cassava production data sourced from FAO, African governments, and IITA's Collaborative Study on Cassava in Africa	No attribution in study, but IITA efforts comprised 80% of total expenditures	World market prices for cassava were utilised for scenario used in the present study	5.10	Significantly demonstrated

GLOSSARY OF ACRONYMS

B-C	Benefit-Cost
CGIAR	Consultative Group on International Agricultural Research
CIAT	International Centre for Tropical Agriculture
CIMMYT	International Maize and Wheat Improvement Centre
CIP	International Potato Centre
GR	Green Revolution
IA	Ex-post Impact Assessment
IAEG	Impact Assessment and Evaluation Group
IARC	International Agricultural Research Centre
ICARDA	International Centre for Agriculture Research in Dry Areas
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IFPRI	International Food Policy Research Institute
IITA	International Institute of Tropical Agriculture
IRRI	International Rice Research Institute
MV	Modern Variety
NARS	National Agricultural Research System
NPV	Net Present Value
SPIA	Standing Panel on Impact Assessment
WARDA	West Africa Rice Development Association